

# Information Nonanticipative Rate Distortion Function and its Applications

Photios A. Stavrou\*, Christos K. Kourtellis†, and Charalambos D. Charalambous\*

## Abstract

This paper investigates applications of nonanticipative Rate Distortion Function (RDF) in a) zero-delay Joint Source-Channel Coding (JSCC) design based on average and excess distortion probability, b) in bounding the Optimal Performance Theoretically Attainable (OPTA) by noncausal and causal codes, and computing the Rate Loss (RL) of zero-delay and causal codes with respect to noncausal codes. These applications are described using two running examples, the Binary Symmetric Markov Source with parameter  $p$ , (BSMS( $p$ )) and the multidimensional partially observed Gaussian-Markov source.

For the multidimensional Gaussian-Markov source with square error distortion, the solution of the nonanticipative RDF is derived, its operational meaning using JSCC design via a noisy coding theorem is shown by providing the optimal encoding-decoding scheme over a vector Gaussian channel, and the RL of causal and zero-delay codes with respect to noncausal codes is computed. For the BSMS( $p$ ) with Hamming distortion, the solution of the nonanticipative RDF is derived, the RL of causal codes with respect to noncausal codes is computed, and an uncoded noisy coding theorem based on excess distortion probability is shown. The information nonanticipative RDF is shown to be equivalent to the nonanticipatory  $\epsilon$ -entropy, which corresponds to the classical RDF with an additional causality or nonanticipative condition imposed on the optimal reproduction conditional distribution.

## Index Terms

Nonanticipative RDF, sources with memory, Joint Source-Channel Coding (JSCC), Binary Symmetric Markov Source (BSMS), multidimensional stationary Gaussian-Markov source, excess distortion probability, bounds.

## I. INTRODUCTION

In lossy compression source coding with fidelity [3], [4], the sequence of real-valued symbols  $x^\infty \triangleq \{x_0, x_1, \dots\}$ ,  $x_j \in \mathcal{X}$ ,  $\forall j \geq 0$ , generated by a source with distribution  $P_{X^\infty}$ , is transformed by the encoder into a sequence of symbols, the compressed representation  $z^\infty \triangleq \{z_0, z_1, \dots\}$  (taking values in a finite alphabet set), which is then transmitted over a noiseless channel. The decoder at the channel output upon observing the compressed representation symbols produces the reproduction sequence  $y^\infty \triangleq \{y_0, y_1, \dots\}$ ,  $y_j \in \mathcal{Y}$ ,  $\forall j \geq 0$ . Such a compression system is called causal [5] if the reproduction symbol  $y_n$  of the source symbol  $x_n$ , depends on the present and past source symbols  $\{x_0, \dots, x_n\}$  but not on the future source symbols  $\{x_{n+1}, x_{n+2}, \dots\}$ , i.e.,  $y_n \triangleq f_n(x_0, \dots, x_n)$ , for some function  $f_n(\cdot)$ ,  $n \geq 0$ . The Optimal Performance Theoretically Attainable (OPTA) by noncausal codes is given by the Rate Distortion Function (RDF) [3], [4], while that of causal codes is given by the entropy rate of the reproduction sequence [5]. Zero-delay source codes are a sub-class of causal codes, with the additional constraint that the compressed representation symbol  $z_n$ , depends on the past and present source symbols  $x^n \triangleq \{x_0, x_1, \dots, x_n\}$ , while the reproduction at the decoder  $y_n$  of the present source symbol  $x_n$ , depends only on the compressed representation  $z^n \triangleq \{z_0, z_1, \dots, z_n\}$ , i.e.,  $z_n = h_n(\{x_j : j = 0, 1, \dots, n\})$  and  $y_n = f_n(\{z_j : j = 0, 1, \dots, n\})$ ,  $\forall n \geq 0$  [6]–[10].

Joint Source-Channel Coding (JSCC) design based on nonanticipative processing (i.e., the encoder-channel-decoder mappings at each time instant do not depend on future symbols), is perhaps, the most efficient zero-delay coding system, in the sense of optimal performance of matching the source characteristics to the channel characteristics, coded or uncoded [11]–[13]. Two such fascinating examples of matching the RDF of an IID source to the capacity of a memoryless channel are the following. *Example-IID-BSS*: The Independent Identically Distributed (IID) Binary Source (BS) with Hamming distortion transmitted uncoded over a symmetric memoryless channel (the distortion and channel parameter are made equal);

*Example-IID-GS*: The IID Gaussian Source (GS) with average squared-error distortion, transmitted over a memoryless Additive White Gaussian Noise (AWGN) channel without feedback, with the encoder and decoder scaling their inputs [13].

In both examples, symbol-by-symbol transmission is optimal, that is, the optimal code rate is “1” in terms of achieving average end-to-end distortion. These examples demonstrate the simplicity of the JSCC design often called probabilistic matching, in operating optimally with zero-delay, in complexity, when this is compared to the asymptotic performance of optimally separating the encoder/decoder to the source and channel encoders/decoders which may cause long processing delays. Recently, *Example-IID-BSS* and *Example-IID-GS* are revisited in [14], using symbol-by-symbol (zero-delay) codes.

This work was financially supported by a medium size University of Cyprus grant entitled “DIMITRIS” and by QNRF, a member of Qatar Foundation, under the project NPRP 6-784-2-329. Part of the material in this paper was presented in IEEE International Symposium on Information Theory, 2014 [1] and in book series “Lecture Notes in Control and Information Sciences” [2].

\* Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus † Department of Electrical and Computer Engineering, Texas A&M University at Qatar Email: {stavrou.fotios, chadcha}@ucy.ac.cy, c.kourtellis@qatar.tamu.edu

In general, very little is known about JSCC design or probabilistic matching based on nonanticipative processing, that of matching the source characteristics to the channel characteristics, as in *Example-IID-BSS* and in *Example-IID-GS*, because the RDF of sources with memory is generally not known. Moreover, only bounds are known on the OPTA by causal and zero-delay codes [12]. For the latter, these bounds are often introduced to quantify the Rate Loss (RL) due to causality and zero-delay of the coding systems compared to that of the noncausal coding systems. Such bounds are elaborated recently in [15], for Gaussian sources with square-error distortion function.

In this paper, we consider the information nonanticipative RDF [16], [17], a variant of the classical information RDF, and we describe its applications in

- (a) zero-delay JSCC design or probabilistic matching of the source to the channel, with respect to nonanticipative processing (zero delay or symbol-by-symbol codes), based on average and excess distortion probability;
- (b) bounding the OPTA by noncausal and causal codes, and computing the RL of zero-delay and causal codes with respect to noncausal codes.

For (a), the design of encoder and decoder for a given {source, channel} pair with respect to a given {distortion function, transmission cost function} pair is constructive. Given a source, first we compute the nonanticipative RDF. Second, we realize the nonanticipative RDF (see Fig. 1.1) with respect to the channel by providing an {encoder, decoder} pair, so that the end-to-end average distortion is achieved, the channel operates at its capacity, and the nonanticipative RDF is equal to the channel capacity. For the multidimensional (vector) stationary Gaussian source with memory and square-error distortion, we show that the solution of the nonanticipative RDF is equivalent to JSCC design of a vector AWGN channel with and without feedback encoding in which the encoder and the decoder are linear. For the special case of IID Gaussian sequence, our JSCC design reduces to the JSCC design of *Example IID-GS* [11], [13]. Moreover, from the solution of the nonanticipative RDF we recover the Schalkwijk-Kailath's coding scheme [18].

For (b), we use the nonanticipative RDF to compute the RL or gap between the OPTA by causal [5] and zero-delay codes with respect to the OPTA by noncausal codes, for both finite alphabet and continuous alphabet valued sources.

Moreover, we show equivalence of the information nonanticipative RDF and its rate to Gorbunov and Pinsker nonanticipatory  $\epsilon$ -entropy and message generation rates [19]–[21], which corresponds to Shannon information RDF with an additional causality constraint imposed on the optimal reproduction distribution. Gorbunov and Pinsker [19] appear to be the first who recognized the importance of nonanticipative RDF in real-time applications [19, p. 1, lines 4-5].

Next, we introduce the nonanticipative RDF, classical RDF and Gorbunov and Pinsker's nonanticipatory  $\epsilon$ -entropy to clarify certain relations between them.

**Nonanticipative RDF.** The information nonanticipative RDF is defined as follows. Consider a source distribution  $P_{X^n}(dx^n)$ , a causal sequence of reproduction distributions  $\{P_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i) : i = 0, 1, \dots, n\}$ , a measurable distortion function  $d_{0,n}(x^n, y^n) : \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \mapsto [0, \infty]$  and an average fidelity set<sup>1</sup>

$$\begin{aligned} \vec{\mathcal{Q}}_{0,n}(D) \triangleq & \left\{ \vec{P}_{Y^n|X^n}(dy^n|x^n) \triangleq \otimes_{i=0}^n P_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i) : \right. \\ & \left. \frac{1}{n+1} \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) (\vec{P}_{Y^n|X^n} \otimes P_{X^n})(dx^n, dy^n) \leq D \right\}. \end{aligned} \quad (1.1)$$

The information nonanticipative RDF is defined by

$$\begin{aligned} R_{0,n}^{na}(D) \triangleq & \inf_{\vec{P}_{Y^n|X^n}(\cdot|x^n) \in \vec{\mathcal{Q}}_{0,n}(D)} \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \log \left( \frac{\vec{P}_{Y^n|X^n}(dy^n|x^n)}{P_{Y^n}(dy^n)} \right) (\vec{P}_{Y^n|X^n} \otimes P_{X^n})(dx^n, dy^n) \\ = & \inf_{\vec{P}_{Y^n|X^n}(\cdot|x^n) \in \vec{\mathcal{Q}}_{0,n}(D)} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{P}_{Y^n|X^n}). \end{aligned} \quad (1.2)$$

Here,  $\mathbb{I}_{X^n \rightarrow Y^n}(\cdot, \cdot)$  is used to denote the functional dependence of  $R_{0,n}^{na}(D)$  on the two distributions  $\{P_{X^n}, \vec{P}_{Y^n|X^n}\}$ . The information nonanticipative RDF rate is defined by

$$R^{na}(D) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na}(D). \quad (1.3)$$

**Classical RDF.** The classical information RDF (often called OPTA by noncausal codes) [3], [4], [22] is defined by

$$R(D) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}(D), \quad R_{0,n}(D) \triangleq \inf_{P_{Y^n|X^n}(\cdot|x^n) \in \mathcal{Q}_{0,n}(D)} I(X^n; Y^n) \quad (1.4)$$

where the fidelity set of reproduction conditional distributions is defined by

$$\mathcal{Q}_{0,n}(D) \triangleq \left\{ P_{Y^n|X^n}(dy^n|x^n) : \frac{1}{n+1} \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) (P_{Y^n|X^n} \otimes P_{X^n})(dx^n, dy^n) \leq D \right\}. \quad (1.5)$$

<sup>1</sup>  $\otimes$  denotes convolution of distributions.

Goblick in [13] applied the  $R(D)$  of an IID Gaussian distributed  $N(0; \sigma_X^2)$  sequence with  $R(D) = \frac{1}{2} \log(\frac{\sigma_X^2}{D})$ ,  $0 \leq D \leq \sigma_X^2$  to provide the JSCC design over a memoryless AWGN channel without feedback. The extension of Goblick's JSCC design to a memoryless AWGN channel with feedback is treated in [23], and makes use of the Schalkwijk-Kailath's coding scheme.

**Nonanticipatory  $\epsilon$ -Entropy.** Gorbunov and Pinsker in [19] introduced the so-called nonanticipatory  $\epsilon$ -entropy,  $R_{0,n}^\epsilon(D)$ , and message generation rate  $R^\epsilon(D)$  as follows

$$R^\epsilon(D) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^\epsilon(D), \quad (I.6)$$

$$R_{0,n}^\epsilon(D) \triangleq \inf_{\substack{P_{Y^n|X^n}(\cdot|x^n) \in \mathcal{Q}_{0,n}(D) \\ X_{i+1}^n \leftrightarrow X^i \leftrightarrow Y^i, \quad i=0,1,\dots,n-1}} I(X^n; Y^n), \quad \forall n \geq 0. \quad (I.7)$$

The same authors in [20], [21], showed that for scalar stationary Gaussian sources (using power spectral densities) that  $R^\epsilon(D)$  tends to Shannon's RDF  $R(D)$ , as  $D \rightarrow 0$ .

We show in Theorem III.3 that  $R_{0,n}^{na}(D) = R_{0,n}^\epsilon(D)$ . The fundamental difference between  $R_{0,n}^{na}(D)$  or  $R_{0,n}^\epsilon(D)$  and  $R_{0,n}(D)$  is the following. For sources with memory, the optimal reproduction distribution of the information nonanticipative RDF,  $R_{0,n}^{na}(D)$  at each time  $i = 0, \dots, n$ , is a causal (or nonanticipative) sequence of conditional distributions  $P_{Y_i|Y^{i-1}, X^i}^*(dy_i|y^{i-1}, x^i)$ , that is, it depend only on the past and present source symbols and past reproduction symbols  $\{x^i, y^{i-1}\}$ . On the other hand, the optimal reproduction distribution of the classical RDF,  $R_{0,n}(D)$  at each time  $i = 0, \dots, n$ , is a sequence of noncausal conditional distributions  $P_{Y_i|Y^{i-1}, X^n}^*(dy_i|y^{i-1}, x^n)$ , which depends not only on  $\{x^i, y^{i-1}\}$  but also on future source symbols  $\{x_{i+1}, \dots, x_n\}$ . For independent sources,  $R_{0,n}^{na}(D)$  and  $R_{0,n}(D)$  coincide.

Next, we discuss certain limitations of the classical information RDF  $R(D)$  with respect to applications, which motivated our interest to investigate the information nonanticipative RDF (I.2).

#### A. Motivation for Nonanticipative RDF

The first limitation of the classical information RDF is the lack of examples of a source with memory for which the exact expression of  $R_{0,n}(D)$  for finite  $n$ , and that of  $R(D) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}(D)$ , are computed, aside from memoryless or Gaussian sources. For example, the RDF of the BSMS( $p$ ) with single letter Hamming distortion function is currently unknown (see Section IV-B for more discussion).

The second limitation of the classical information RDF  $R_{0,n}(D)$  is the anticipative (noncausal) nature of the optimal reproduction distributions  $\{P_{Y_i|Y^{i-1}, X^n}^*(dy_i|y^{i-1}, x^n) : i = 0, 1, \dots, n\}$ , because they depend on future source symbols even for single letter distortion functions  $d_{0,n}(x^n, y^n) \triangleq \sum_{i=0}^n \rho(x_i, y_i)$ . This anticipative structure of the optimal reproduction distribution corresponding to  $R_{0,n}(D)$  implies that, in general, the classical information RDF cannot be used in JSCC design using nonanticipative processing or uncoded transmission (i.e., the {encoder, decoder} pair processes at each time instant source symbol without dependence on future source symbols), also called probabilistic matching of the source to the channel [12], unless the source is memoryless, such as, the *Example-IID-BS* and the *Example-IID-GS* discussed earlier [11], [14], [24], [25]. Indeed, to perform the JSCC design using nonanticipative processing it is necessary that 1) the exact expression of the RDF of the source is known or computed, and 2) its corresponding optimal reproduction distribution is realizable using nonanticipative processing via {encoder, channel, decoder} mappings processing symbols without dependence on future symbols, as shown in Fig. I.1. This means that for general sources with memory, the optimal conditional reproduction distribution of  $R_{0,n}(D)$  should satisfy the Markov chain (MC) (i.e., conditional independence)

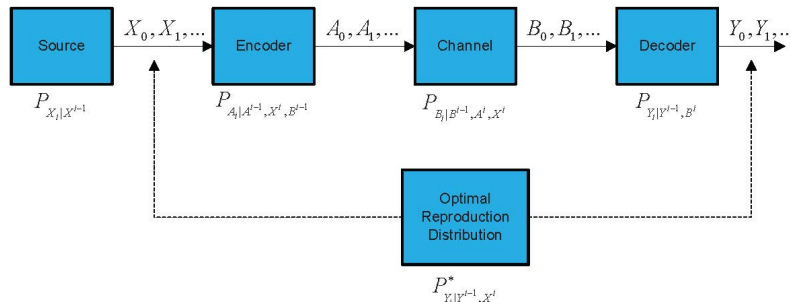


Fig. I.1: Probabilistic matching based on nonanticipative transmission.

$$P_{Y_i|Y^{i-1},X^n}^*(dy_i|y^{i-1},x^\infty) = P_{Y_i|Y^{i-1},X^i}^*(dy_i|y^{i-1},x^i) \iff X_{i+1}^\infty \leftrightarrow (X^i, Y^{i-1}) \leftrightarrow Y_i, \quad i = 0, 1, \dots \quad (\text{I.8})$$

Alternatively stated, if (I.8) does not hold, then nonanticipative processing of information in the sense of Fig. I.1 is violated.

### B. Summary of Main Results and Related Literature

Next, we present a summary of the contributions in this paper and we discuss relations to existing literature.

**Relations to Nonanticipatory  $\epsilon$ -Entropy.** In subsection III-A we show that the information nonanticipative RDF,  $R_{0,n}^{na}(D)$  and its rate  $R^{na}(D)$  (see (I.2), (I.3)), are equivalent to Gorbunov and Pinsker's nonanticipatory  $\epsilon$ -entropy and message generation rate [19],  $R_{0,n}^\epsilon(D)$  and  $R^\epsilon(D)$ , respectively. In subsection III-B, we combine the existence of solution to the nonanticipative RDF shown in [16, Theorem III.4], and the main theorems in [19] to establish that for general stationary sources, the limit exists, and the limit and infimum operations can be interchanged, that is,

$$\begin{aligned} R^{na}(D) &= \lim_{n \rightarrow \infty} \inf_{\vec{P}_{Y^n|X^n} \in \vec{\mathcal{Q}}_{0,n}(D)} \frac{1}{n+1} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{P}_{Y^n|X^n}) \\ &= \inf_{\vec{P}_{Y^\infty|X^\infty} \in \vec{\mathcal{Q}}_{0,\infty}(D)} \lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{P}_{Y^n|X^n}) \equiv \vec{R}^{na}(D) < \infty. \end{aligned} \quad (\text{I.9})$$

and that the optimal reproduction distribution of  $R^{na}(D)$  corresponds to jointly stationary source-reproduction pair  $\{(X_i, Y_i) : i = 0, 1, \dots\}$ .

**Closed Form Expression of Information Nonanticipative RDF (Stationary Solution), Properties, and Examples.** In Section IV-A, we characterize the stationary solution of the nonanticipative RDF and some of its properties. One of the properties is analogous to the Shannon lower bound of the classical RDF [3]. In subsections IV-B and IV-C, we use these to compute the information theoretic nonanticipative RDF in closed form for the following two running examples.

*Example 1: BSMS( $p$ ) with single letter Hamming distortion.* For the Binary Symmetric Markov Source with parameter  $p$  (BSMS( $p$ )) and single letter Hamming distortion, we show that

$$R^{na}(D) = \begin{cases} H(m) - H(D) & \text{if } D \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}, \quad m = 1 - p - D + 2pD. \quad (\text{I.10})$$

Note that, for  $p = \frac{1}{2}$ , then  $R^{na}(D)$  reduces to the classical RDF,  $R(D)$ , of IID memoryless source, as expected. Recently,  $R_{0,n}^{na}(D)$  of BSMS( $p$ ) is utilized in [26] to demonstrate an example of an opportunistic scheduling system. In section VI we use (I.10) to obtain tighter lower bounds on the OPTA by causal and noncausal codes, compared to those given in [27].

*Example 2: Multidimensional Gaussian-Markov source with square-error distortion.* For the multidimensional partially observed stationary Gauss-Markov source, described in state space form (this includes autoregressive models) by

$$\begin{cases} Z_{t+1} = AZ_t + BW_t, \quad Z_0 \sim N(0, \bar{\Sigma}_0), \quad t = 0, 1, \dots \\ X_t = CZ_t + NV_t, \quad t = 0, 1, \dots \end{cases} \quad (\text{I.11})$$

where  $Z_t \in \mathbb{R}^m$  is the state (unobserved) process driven by a multidimensional Gaussian noise process  $\{W_t : t = 0, 1, \dots\}$ , and

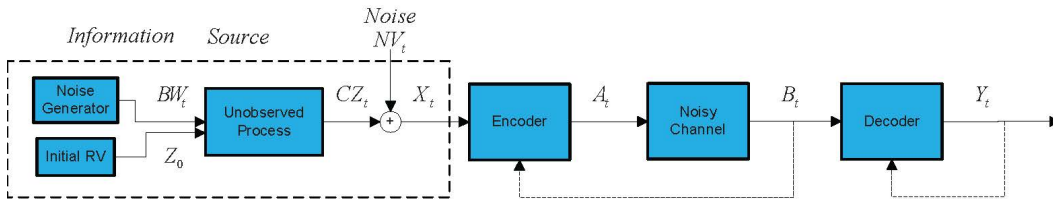


Fig. I.2: JSCC design system of partially observed multidimensional Gaussian-Markov source.

$X_t \in \mathbb{R}^p$  is the observed source process corrupted by a multidimensional additive Gaussian noise process  $\{V_t : t = 0, 1, \dots\}$  (see<sup>2</sup> Fig. I.2), we utilize the characterization of the solution of  $R_{0,n}^{na}(D)$  obtained in Theorem IV.7, to show that for a square error distortion function, the nonanticipative RDF is given by

$$R^{na}(D) = \frac{1}{2} \sum_{i=0}^p \log \left( \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} \right) = \frac{1}{2} \log \frac{|\Lambda_\infty|}{|\Delta_\infty|} \quad (\text{I.12})$$

<sup>2</sup>In this application, the objective is to reconstruct  $\{X_t : t = 0, 1, \dots\}$  by  $\{Y_t : t = 0, 1, \dots\}$  with respect to a certain fidelity of reproduction. Under certain assumptions [28]  $\{X_t : t = 0, 1, \dots, n\}$  is ergodic (although the  $A$  matrix of  $\{Z_t : t = 0, 1, \dots, n\}$  may have unstable eigenvalues).

where

$$\Lambda_\infty \triangleq \lim_{n \rightarrow \infty} \Lambda_t, \Lambda_t \triangleq \mathbb{E} \left\{ \left( X_t - \mathbb{E}(X_t | Y^{t-1}) \right) \left( X_t - \mathbb{E}(X_t | Y^{t-1}) \right)^{tr} \right\} \equiv (\text{IV.21}), \Delta_\infty = \text{diag}\{\delta_{\infty,1}, \dots, \delta_{\infty,p}\}$$

Here,  $\{\lambda_{\infty,1}, \dots, \lambda_{\infty,p}\}$  are the steady state eigenvalues of  $\Lambda_\infty$ , and

$$\delta_{\infty,i} \triangleq \begin{cases} \xi_\infty & \text{if } \xi_\infty \leq \lambda_{\infty,i} \\ \lambda_{\infty,i} & \text{if } \xi_\infty > \lambda_{\infty,i} \end{cases}, i = 1, \dots, p, \sum_{i=1}^p \delta_{\infty,i} = D.$$

In addition, in subsection IV-C, we recover from (I.12), several special cases. These include the expression of the nonanticipative RDF for the *scalar stationary fully observed Gaussian-Markov source* (i.e., corresponding to  $p = 1, C = 1, N = 0, A = \alpha, B = \sigma_W$ ) also obtained in [15, Theorem 3] via alternative methods, and for *IID Gaussian sources*  $\{X_t : t = 0, 1, \dots\} \sim N(0; \sigma_X^2)$  the known expression  $R(D) = \frac{1}{2} \log \frac{\sigma_X^2}{D}, \sigma_X^2 \geq D \geq 0$ .

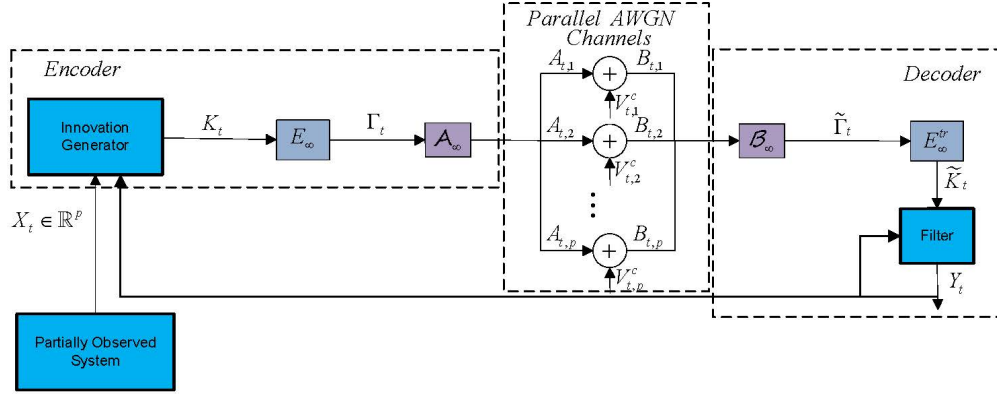


Fig. I.3: Realization of the optimal reproduction distribution:  $K_t = X_t - \mathbb{E}\{X_t | Y^{t-1}\}$ ,  $\tilde{K}_t = Y_t - \mathbb{E}\{X_t | Y^{t-1}\}$ ,  $\{E_\infty, E_\infty^{tr}\}$ =unitary transformations,  $\{A_\infty, B_\infty\}$ =scaling matrices,  $\{\Gamma_t, \tilde{\Gamma}_t\}$ = scaling matrices.

We note that recently in [29] the importance of example (I.11) is elaborated for the special case of fully observed Gaussian-Markov sources, i.e.,  $N = 0, C = I$  (i.e.,  $X_t = Z_t$ ) in the context of the so-called sequential rate distortion theory [8], which is equivalent to the nonanticipative RDF. The authors in [29] provide an approximation solution by utilizing a semidefinite programming approach, while we give the exact closed form expression (I.12) for the more general model (I.11).

**Joint Source-Channel Coding: Symbol-by-Symbol Transmission.** In Section V, we show that (I.12) is achievable by the JSCC design system depicted in Fig. I.3, where the encoder operates using symbol-by-symbol codes, achieves channel capacity  $C(P)$  ( $P$  is the power allocated for transmission), while the end-to-end average distortion is met, and  $R^{na}(D) = C(P)$ .

Moreover, we demonstrate that the general JSCC design system of Fig. I.3 gives as degraded cases (corresponding to  $p = 1, C = 1, N = 0, A = \alpha, B = \sigma_W$ ), the following new examples of JSCC designs operating based on symbol-by-symbol codes.

- (i) The optimal coding scheme of a *scalar Gaussian-Markov Source* over a *memoryless AWGN with feedback* (see Fig. I.4);
- (ii) the optimal coding scheme of a *scalar Gaussian-Markov Source* over a *memoryless AWGN channel without feedback* (see Fig. I.5).

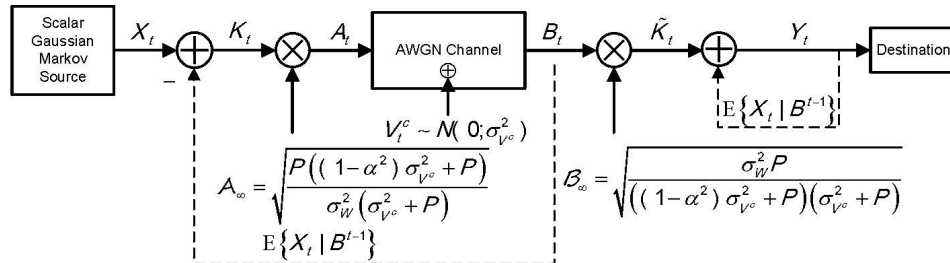


Fig. I.4: Reduction of Fig. I.3 to  $p = 1, C = 1, N = 0, A = \alpha, |\alpha| < 1, B = \sigma_W$ , which corresponds to the JSCC design system of a scalar Gaussian Markov Source given by (IV.35) over a memoryless AWGN channel with feedback,  $K_t = X_t - \mathbb{E}\{X_t | B^{t-1}\}$ .

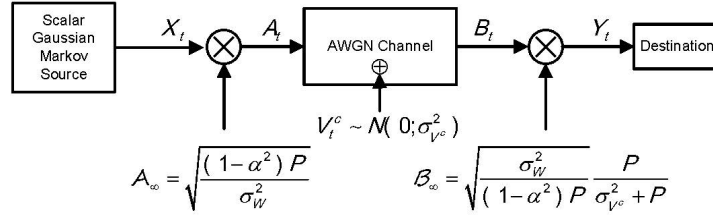


Fig. I.5: Reduction of Fig. I.3 to  $p = 1$ ,  $C = 1$ ,  $N = 0$ ,  $A = \alpha$ ,  $|\alpha| < 1$ ,  $B = \sigma_W$ , which corresponds to the JSCC design system of a scalar Gaussian Markov source given by (IV.35) over a memoryless AWGN channel without feedback.

Note that Fig. I.5 with  $\alpha = 0$ , reduces to Goblick's JSCC design system of the *scalar IID Gaussian process transmitted over a memoryless AWGN channel without feedback*. Finally, we demonstrate how the realization scheme of Fig. I.3 recovers the Schalkwijk-Kailath's coding scheme which achieves the capacity of a memoryless AWGN channel with feedback [18].

When  $\{X_t : t = 0, 1, \dots, n\}$  is a vector with independent components  $\{X_t^1 : t = 0, 1, \dots, n\}, \dots, \{X_t^n : t = 0, 1, \dots, n\}$  then the analogues of Fig. I.4, Fig. I.5 are easily obtained, and involve JSCC design of a vector source over a vector AWGN channel.

**Bounds on Noncausal and Causal Codes.** In Section VI we use  $R^{na}(D)$  to derive the following bounds for  $R(D)$ , and the OPTA by causal codes [5], denoted by  $r^{c,+}(D)$ :

$$r^{c,+}(D) \geq R^{na}(D) \geq R(D), \quad R_{0,n}^{na}(D) \geq R_{0,n}(D), \quad \forall n \geq 0. \quad (\text{I.13})$$

Based on these bounds we evaluate the RL due to causality by using  $R^{na}(D)$ . This part compliments previous work by Linder and Zamir [30], who showed using the results in [5], that at high resolution (small distortion), an optimal causal code for stationary source with finite differential entropy and square error distortion, consists of uniform quantizers followed by a sequence of entropy coders, and that the RL due to causality is given by the so-called space-filling loss of quantizers, which is at most  $\frac{1}{2} \ln \left( \frac{2\pi e}{12} \right) \simeq 0.254$  bits/sample.

For arbitrary Gaussian stationary sources with memory defined by (I.11) and square-error distortion, we use the explicit expression of  $R^{na}(D)$  (i.e., (I.12)) in (I.13) to evaluate the RL due to causality. Our analysis of vector Gaussian stationary sources (I.11) compliment Gorbunov and Pinsker computation in [20], [21], where it is shown (for scalar Gaussian sources), using power spectral densities that  $R^\varepsilon(D)$  tends to Shannon's RDF  $R(D)$ , as  $D \rightarrow 0$ .

Moreover, for stationary Gaussian processes with square-error distortion, our results compliment recent contributions obtained in [15], on the gap between the OPTA by causal codes, nonanticipatory  $\epsilon$ -entropy,  $R^\varepsilon(D)$ , and classical RDF,  $R(D)$ . Specifically, in [15] it is shown that for zero-mean Gaussian sources with square-error distortion (and bounded differential entropy rate), the OPTA by causal codes exceeds  $R^\varepsilon(D)$  by less than approximately 0.254 bits/sample. The analysis in [15] includes a closed-form expression for  $R^\varepsilon(D)$  only when the source is first-order Markov with square-error distortion, while for general Gaussian sources it is shown that  $r^{c,+}(D) \leq R^\varepsilon(D) + \frac{1}{2} \log_2(2\pi e)$  bits/sample, but no expression is given for  $R^\varepsilon(D)$ . Since we show  $R^\varepsilon(D) = R^{na}(D)$ , then we can use this equivalence to evaluate the gap between the OPTA by causal codes and  $R(D)$ , for arbitrary multidimensional stationary Gaussian sources (partially observed), via the bound  $R^{na}(D) - R(D) \leq r^{c,+}(D) - R(D)$ . Note that the RL due to zero-delay and noncausal codes does not exceed  $\frac{1}{2} \log \frac{|\Lambda_\infty|}{|\Delta_\infty|} - R(D)$ , where  $R(D)$  can be computed using power spectral densities [3].

For the case of BSMS( $p$ ), we show that the RL due to causality, i.e.,  $R^{na}(D) - R(D)$ , does not exceed  $H(m) - H(p)$ , for the region  $0 \leq D \leq D_c$ ,  $m = 1 - p - D + 2pD$ , while for the region  $D_c \leq D \leq \frac{1}{2}$ , we compare the upper bound given in [27] with the upper bound obtained via  $R^{na}(D)$ . Our simulations and properties of  $R^{na}(D)$  illustrate that the proposed upper bound on  $R(D)$  based on  $R^{na}(D)$  is more reliable compared to the one given in [27].

**Noisy and Noiseless Coding Theorems.** In Section VII-A, we show achievability of nonanticipative RDF using the JSCC design system (Fig. I.1) based on nonanticipative transmission, with respect to the excess distortion probability between the source symbols and their reproductions. That is, we show achievability based on probabilistic matching of the source and the channel with respect to excess distortion probability. For the multidimensional stationary Gaussian source, we use the JSCC design of Fig. I.3, and we apply Chernoff's bound to compute the error exponent of the excess distortion probability. For the BSMS( $p$ ), we apply uncoded transmission, and we invoke Hoeffding's inequality and large deviations theory to compute the error exponent of the excess distortion probability. In Section VII-B, we briefly point out that the classical noiseless coding theorem derived in [8, Chapter 5] for two dimensional sources (such as video coding applications) with per-sample or average distortion function is applicable, giving an alternative operational meaning to  $R_{0,n}^{na}(D)$ .

## II. INFORMATION NONANTICIPATIVE RDF

In this section, we introduce the definition of information nonanticipative RDF, for general source/reproduction alphabets modelled by complete separable metric spaces (Polish spaces) so that our analysis holds for both finite and continuous alphabets (because we present examples for both). Then we state some properties of information nonanticipative RDF which follow directly from the general properties of directed information found in [31].

### A. Information Nonanticipative RDF

Let  $\mathbb{N} \triangleq \{0, 1, 2, \dots\}$ , and  $\mathbb{N}^n \triangleq \{0, 1, 2, \dots, n\}$ . Introduce the source spaces  $\{(\mathcal{X}_n, \mathcal{B}(\mathcal{X}_n)) : n \in \mathbb{N}\}$  and the reproduction spaces  $\{(\mathcal{Y}_n, \mathcal{B}(\mathcal{Y}_n)) : n \in \mathbb{N}\}$ , where  $\mathcal{X}_n, \mathcal{Y}_n, n \in \mathbb{N}$  are Polish spaces, and  $\mathcal{B}(\mathcal{X}_n)$  and  $\mathcal{B}(\mathcal{Y}_n)$  are Borel  $\sigma$ -algebras of subsets of  $\mathcal{X}_n$  and  $\mathcal{Y}_n$ , respectively. Points in  $\mathcal{X}^{\mathbb{N}} \triangleq \times_{n \in \mathbb{N}} \mathcal{X}_n$ ,  $\mathcal{Y}^{\mathbb{N}} \triangleq \times_{n \in \mathbb{N}} \mathcal{Y}_n$  are denoted by  $\mathbf{x} \triangleq \{x_0, x_1, \dots\} \in \mathcal{X}^{\mathbb{N}}$ ,  $\mathbf{y} \triangleq \{y_0, y_1, \dots\} \in \mathcal{Y}^{\mathbb{N}}$ , respectively, while their restrictions to finite coordinates for any  $n \in \mathbb{N}$  are denoted by  $x^n \triangleq \{x_0, x_1, \dots, x_n\} \in \mathcal{X}_{0,n}$ ,  $y^n \triangleq \{y_0, y_1, \dots, y_n\} \in \mathcal{Y}_{0,n}$ . Let  $\mathcal{B}(\mathcal{X}^{\mathbb{N}}) \triangleq \odot_{i \in \mathbb{N}} \mathcal{B}(\mathcal{X}_i)$  denote the  $\sigma$ -algebra on  $\mathcal{X}^{\mathbb{N}}$  generated by cylinder sets  $\{\mathbf{x} = (x_0, x_1, \dots) \in \mathcal{X}^{\mathbb{N}} : x_0 \in A_0, x_1 \in A_1, \dots, x_n \in A_n\}$ ,  $A_i \in \mathcal{B}(\mathcal{X}_i)$ ,  $i = 0, 1, \dots, n, n \in \mathbb{N}$ , and similarly for  $\mathcal{B}(\mathcal{Y}^{\mathbb{N}}) \triangleq \odot_{i \in \mathbb{N}} \mathcal{B}(\mathcal{Y}_i)$ . Thus,  $\mathcal{B}(\mathcal{X}_{0,n})$  and  $\mathcal{B}(\mathcal{Y}_{0,n})$  denote the  $\sigma$ -algebras of cylinder sets in  $\mathcal{X}^{\mathbb{N}}$  and  $\mathcal{Y}^{\mathbb{N}}$ , respectively, with bases over  $A_i \in \mathcal{B}(\mathcal{X}_i)$ , and  $B_i \in \mathcal{B}(\mathcal{Y}_i)$ ,  $i = 0, 1, \dots, n$ , respectively. The set of probability distributions on any measurable space  $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$  is denoted by  $\mathcal{M}(\mathcal{X})$ .

**Source Distribution.** The source is a collection of conditional probability distributions  $\{P_{X_n|X^{n-1}}(dx_n|x^{n-1}) : n \in \mathbb{N}\}$ . For a cylinder set  $B \triangleq \{\mathbf{x} \in \mathcal{X}^{\mathbb{N}} : x_0 \in B_0, x_1 \in B_1, \dots, x_n \in B_n\}$ ,  $B_i \in \mathcal{B}(\mathcal{X}_i)$ ,  $i = 0, \dots, n$ , we can define the probability distribution  $P_{X^n}(B_{0,n})$  on  $\mathcal{B}(\mathcal{X}_{0,n})$  by

$$P_{X^n}(B_{0,n}) \triangleq \int_{B_0} P_{X_0}(dx_0) \dots \int_{B_n} P_{X_n|X^{n-1}}(dx_n|x^{n-1}), \quad B_{0,n} = \times_{i=0}^n B_i. \quad (\text{II.1})$$

For each  $n \in \mathbb{N}$ , we use the notation  $\mathcal{Q}_n(\mathcal{X}_n|\mathcal{X}_{0,n-1}) \triangleq \{P_{X_n|X^{n-1}}(\cdot|x^{n-1}) \in \mathcal{M}(\mathcal{X}_n) : x^{n-1} \in \mathcal{X}_{0,n-1}\}$  to emphasize that  $P_{X_n|X^{n-1}}(\cdot|x^{n-1})$  is a probability distribution on  $\mathcal{X}_n$  for fixed  $x^{n-1} \in \mathcal{X}_{0,n-1}$  and  $P_{X_n|X^{n-1}}(dx_n|\cdot)$  is a measurable function on  $x^{n-1} \in \mathcal{X}_{0,n-1}$  for fixed  $dx_n$ .

**Reproduction Distribution.** The reproduction distribution is a collection of conditional probability distributions  $\{Q_{Y_n|Y^{n-1}, X^n}(dy_n|y^{n-1}, x^n) : n \in \mathbb{N}\}$ . By our notation, for each  $n \in \mathbb{N}$ ,  $Q_{Y_n|Y^{n-1}, X^n}(\cdot|\cdot, \cdot) \in \mathcal{Q}_n(\mathcal{Y}_n|\mathcal{Y}_{0,n-1} \times \mathcal{X}_{0,n})$ . For a cylinder set  $C \triangleq \{\mathbf{y} \in \mathcal{Y}^{\mathbb{N}} : y_0 \in C_0, y_1 \in C_1, \dots, y_n \in C_n\}$ ,  $C_i \in \mathcal{B}(\mathcal{Y}_i)$ ,  $i = 0, \dots, n$ ,  $C_{0,n} = \times_{i=0}^n C_i$ , we define the family of conditional probability distributions  $\vec{Q}_{Y^n|X^n}(C_{0,n}|x^n)$  on  $\mathcal{B}(\mathcal{Y}_{0,n})$  by

$$\vec{Q}_{Y^n|X^n}(C_{0,n}|x^n) \triangleq \int_{C_0} Q_{Y_0|X_0}(dy_0|x_0) \dots \int_{C_n} Q_{Y_n|Y^{n-1}, X^n}(dy_n|y^{n-1}, x^n). \quad (\text{II.2})$$

Throughout the paper, we assume with abuse of notation that  $Q_{Y_0|X_0}(dy_0|x_0) \equiv Q_{Y_0|X_0, Y^{-1}}(dy_0|x_0, y^{-1})$  and that either  $P_{Y^{-1}}(dy^{-1})$  is fixed or  $Y^{-1} = y^{-1}$  is fixed.

We denote the family of probability distributions (II.2) by the set

$$\vec{\mathcal{Q}}(\mathcal{Y}_{0,n}|\mathcal{X}_{0,n}) \triangleq \left\{ \vec{Q}_{Y^n|X^n}(\cdot|x^n) \in \mathcal{M}(\mathcal{Y}_{0,n}) : x^n \in \mathcal{X}_{0,n}, \vec{Q}_{Y^n|X^n}(\cdot|x^n) \text{ is defined by (II.2)} \right\}.$$

Thus, any element  $\vec{Q}_{Y^n|X^n}(\cdot|\cdot) \in \vec{\mathcal{Q}}(\mathcal{Y}_{0,n}|\mathcal{X}_{0,n})$  is represented by (II.2).

The following theorem is derived in [31].

#### Theorem II.1. [31](Convexity)

The set of probability distributions  $\vec{Q}_{Y^n|X^n}(\cdot|x^n)$  (for a fixed  $x^n \in \mathcal{X}_{0,n}$ ) defined by (II.2) is a convex subset of  $\mathcal{M}(\mathcal{Y}_{0,n})$ .

Theorem II.1 states that for any  $\lambda \in [0, 1]$ , and  $\vec{Q}_{Y^n|X^n}^1(\cdot|x^n)$ ,  $\vec{Q}_{Y^n|X^n}^2(\cdot|x^n)$  two probability measures on  $(\mathcal{Y}_{0,n}, \mathcal{B}(\mathcal{Y}_{0,n}))$  of the form (II.2), then  $\lambda \vec{Q}_{Y^n|X^n}^1(\cdot|x^n) + (1 - \lambda) \vec{Q}_{Y^n|X^n}^2(\cdot|x^n)$  is also a probability measure on  $(\mathcal{Y}_{0,n}, \mathcal{B}(\mathcal{Y}_{0,n}))$  of the form (II.2).

Given the source distribution  $P_{X^n}(\cdot) \in \mathcal{M}(\mathcal{X}_{0,n})$  and reproduction distribution  $\vec{Q}_{Y^n|X^n}(\cdot|\cdot) \in \vec{\mathcal{Q}}(\mathcal{Y}_{0,n}|\mathcal{X}_{0,n})$  define the following measures.

**P1:** The joint distribution on  $\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}$  defined uniquely for  $A_i \in \mathcal{B}(\mathcal{X}_i)$ ,  $B_i \in \mathcal{B}(\mathcal{Y}_i)$ ,  $\forall i \in \mathbb{N}^n$ , by

$$\begin{aligned} P_{X^n, Y^n}(dx^n, dy^n) &\triangleq (P_{X^n} \otimes \vec{Q}_{Y^n|X^n})(\times_{i=0}^n (A_i \times B_i)) \\ &\triangleq \int_{A_0} P_{X_0}(dx_0) \int_{B_0} Q_{Y_0|X_0}(dy_0|x_0) \dots \int_{A_n} P_{X_n|X^{n-1}}(dx_n|x^{n-1}) \int_{B_n} Q_{Y_n|Y^{n-1}, X^n}(dy_n|y^{n-1}, x^n). \end{aligned}$$

**P2:** The marginal distributions on  $\mathcal{Y}_{0,n}$  defined uniquely for  $B_i \in \mathcal{B}(\mathcal{Y}_i)$ ,  $\forall i \in \mathbb{N}^n$ , by

$$P_{Y^n}(\times_{i=0}^n B_i) \triangleq (P_{X^n} \otimes \vec{Q}_{Y^n|X^n})(\times_{i=0}^n (\mathcal{X}_i \times B_i)).$$

**P3:** The product probability distribution  $\vec{\Pi}_{0,n} : \mathcal{B}(\mathcal{X}_{0,n}) \odot \mathcal{B}(\mathcal{Y}_{0,n}) \mapsto [0, 1]$  defined uniquely for  $A_i \in \mathcal{B}(\mathcal{X}_i)$ ,  $B_i \in \mathcal{B}(\mathcal{Y}_i)$ ,  $\forall i \in \mathbb{N}^n$ , by

$$\begin{aligned} \vec{\Pi}_{0,n}(\times_{i=0}^n (A_i \times B_i)) &\triangleq (P_{X^n} \times P_{Y^n})(\times_{i=0}^n (A_i \times B_i)) \\ &= \int_{A_0} P_{X_0}(dx_0) \int_{B_0} P_{Y_0}(dy_0) \dots \int_{A_n} P_{X_n|X^{n-1}}(dx_n|x^{n-1}) \int_{B_n} P_{Y_n|Y^{n-1}}(dy_n|y^{n-1}) \end{aligned}$$

where the notation is  $P_{Y_0}(dy_0) \equiv P_{Y_0|Y^{-1}}(dy_0|y^{-1})$ .

The information theoretic measure associated with nonanticipative RDF is a special case of directed information<sup>3</sup> [8], defined via relative entropy  $\mathbb{D}(\cdot||\cdot)$  as follows.

$$I_{P_{X^n}}(X^n \rightarrow Y^n) \triangleq \mathbb{D}(P_{X^n} \otimes \vec{Q}_{Y^n|X^n} || \vec{\Pi}_{0,n}) \quad (\text{II.3})$$

$$= \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \log \left( \frac{\vec{Q}_{Y^n|X^n}(dy^n|x^n)}{P_{Y^n}(dy^n)} \right) (P_{X^n} \otimes \vec{Q}_{Y^n|X^n})(dx^n, dy^n) \quad (\text{II.4})$$

$$\equiv \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n}). \quad (\text{II.5})$$

The RHS of (II.4) is obtained by using the chain rule of relative entropy [32], [33, Chapter 3]. In (II.5) we use the notation  $\mathbb{I}_{X^n \rightarrow Y^n}(\cdot, \cdot)$  to indicate the functional dependence of  $I_{P_{X^n}}(X^n \rightarrow Y^n)$  on  $\{P_{X^n}, \vec{Q}_{Y^n|X^n}\}$ .

We recall the following convexity result derived in [31] for subsequent use.

**Theorem II.2.** [31] (Convexity of information nonanticipative RDF)

$\mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n})$  is a convex functional of  $\vec{Q}_{Y^n|X^n}(\cdot|x^n) \in \mathcal{M}(\mathcal{Y}_{0,n})$  for a fixed  $P_{X^n}(\cdot) \in \mathcal{M}(\mathcal{X}_{0,n})$ , and a concave functional of  $P_{X^n}(\cdot) \in \mathcal{M}(\mathcal{X}_{0,n})$  for a fixed  $\vec{Q}_{Y^n|X^n}(\cdot|x^n) \in \mathcal{M}(\mathcal{Y}_{0,n})$ .

For each  $n \in \mathbb{N}$ , let  $d_{0,n} : \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \mapsto [0, \infty]$  be a measurable distortion function. The fidelity of reproduction of  $y^n \in \mathcal{Y}_{0,n}$  by  $x^n \in \mathcal{X}_{0,n}$  is defined by the set of conditional distributions

$$\vec{\mathcal{Q}}_{0,n}(D) \triangleq \left\{ \vec{Q}_{Y^n|X^n}(\cdot|x^n) \in \mathcal{M}(\mathcal{Y}_{0,n}) : \frac{1}{n+1} \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) (P_{X^n} \otimes \vec{Q}_{Y^n|X^n})(dx^n, dy^n) \leq D \right\} \quad (\text{II.6})$$

for some  $D \geq 0$ . Denote by  $\vec{\mathcal{Q}}_{0,\infty}(D)$  the corresponding set in (II.6), when the fidelity is replaced by  $\lim_{n \rightarrow \infty} \frac{1}{n+1} \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) (P_{X^n} \otimes \vec{Q}_{Y^n|X^n})(dx^n, dy^n) \leq D$ .

The information nonanticipative RDF is defined as follows.

**Definition II.3.** (Information nonanticipative RDF)

(1) The information nonanticipative RDF is defined by

$$R_{0,n}^{na}(D) \triangleq \inf_{\vec{Q}_{Y^n|X^n}(\cdot|x^n) \in \vec{\mathcal{Q}}_{0,n}(D)} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n}) \quad (\text{II.7})$$

provided the infimum over  $\vec{\mathcal{Q}}_{0,n}(D)$  in (II.7) exists; if not we set  $R_{0,n}^{na}(D) = +\infty$ .

(2) The information nonanticipative RDF rate is defined by

$$R^{na}(D) = \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na}(D) \quad (\text{II.8})$$

provided the limit on the RHS of (II.8) exists; if the infimum over  $\vec{\mathcal{Q}}_{0,n}(D)$  in (II.7) does not exist then we set  $R^{na}(D) = +\infty$ .

In addition, define

$$\vec{R}^{na}(D) \triangleq \inf_{\vec{Q}_{Y^\infty|X^\infty}(\cdot|x^\infty) \in \vec{\mathcal{Q}}_{0,\infty}(D)} \lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n}) \geq R^{na}(D). \quad (\text{II.9})$$

<sup>3</sup>Directed information corresponds to  $\{P_{X_n|X^{n-1}}(dx_i|x^{n-1}) : n \in \mathbb{N}\}$  and  $P_{X^n}(\cdot)$  replaced by  $\{P_{X_n|X^{n-1}, Y^{n-1}}(dx_n|x^{n-1}, y^{n-1}) : n \in \mathbb{N}\}$  and  $\vec{P}_{X^n|Y^{n-1}}(dx^n|y^{n-1}) \triangleq \otimes_{i=0}^n P_{X_i|X^{i-1}, Y^{i-1}}(dx_i|x^{i-1}, y^{i-1})$ , respectively, in the construction of measures **P1-P3**, and (II.3).



Since, in general,  $\overrightarrow{R}^{na}(D) \geq R^{na}(D)$ , then,  $R^{na}(D)$  is more natural than  $\overrightarrow{R}^{na}(D)$ . By analogy with the definition of classical RDF, one may assume that  $\{(X_i, Y_i) : i = 0, 1, \dots\}$  is jointly stationary and ergodic process or  $\frac{1}{n+1} \log \left( \frac{\overrightarrow{Q}_{Y^n|X^n}(\cdot|x^n)}{P_{Y^n}(\cdot)} \right)(y^n)$  is information stable. However, we do not know á priori whether the joint process  $\{(X_i, Y_i) : i = 0, 1, \dots\}$  is stationary.

### III. EQUIVALENCE OF INFORMATION NONANTICIPATIVE RDF AND NONANTICIPATORY $\epsilon$ -ENTROPY

In this section, we first show equivalence of the information nonanticipative RDF (see Definition II.3) to Gorbunov and Pinsker's [19] nonanticipatory  $\epsilon$ -entropy defined by (I.6), (I.7), respectively. Then, we consider consistent stationary sources as defined by Gorbunov and Pinsker [19], and we use certain results from [16], to establish equality of the limiting expressions in (I.9), finiteness of  $R^{na}(D)$ , and that for stationary sources the infimum over  $\overrightarrow{Q}_{Y^n|X^n}(\cdot|x^n) \in \overrightarrow{Q}_{0,n}(D)$  is achieved, and it is realizable by stationary source-reproduction pairs  $\{(X_n, Y_n) : n \in \mathbb{N}\}$ .

#### A. Equivalence of Nonanticipative RDF and Nonanticipatory $\epsilon$ -Entropy

For a given a source  $P_{X^n}(\cdot) \in \mathcal{M}(\mathcal{X}_{0,n})$  and a reproduction  $P_{Y^n|X^n}(\cdot|\cdot) \in \mathcal{Q}_{0,n}(D) \subset \mathcal{Q}(\mathcal{Y}_{0,n}|\mathcal{X}_{0,n})$ , Gorbunov and Pinsker [19] restricted the fidelity set of classical RDF,  $R_{0,n}(D)$ ,  $\mathcal{Q}_{0,n}(D)$  defined by (I.5), to those reproduction distributions which satisfy the following MC.

$$X_{n+1}^\infty \leftrightarrow X^n \leftrightarrow Y^n \iff P_{Y^n|X^\infty}(dy^n|x^\infty) = P_{Y^n|X^n}(dy^n|x^n), \quad n = 0, 1, \dots \quad (\text{III.1})$$

Then they introduced the nonanticipatory  $\epsilon$ -entropy and nonanticipatory message generation rate defined by (I.6), (I.7), respectively. In addition, they defined

$$\overrightarrow{R}^\epsilon(D) \triangleq \inf_{\substack{P_{Y^\infty|X^\infty}(\cdot|x^\infty) \in \mathcal{Q}_{0,\infty}(D) : \\ X_{i+1}^\infty \leftrightarrow X^i \leftrightarrow Y^i, \quad i=0,1,\dots}} \lim_{n \rightarrow \infty} \frac{1}{n+1} I(X^n; Y^n) \geq R^\epsilon(D). \quad (\text{III.2})$$

The MC constraint (III.1) is a probabilistic version of a deterministic causal reproduction coder (cascade of an encoder-decoder (ED)) [5], defined as follows.

**Definition III.1.** [5](Causal reproduction coder)

A reproduction coder  $f_i : \mathcal{X}_{0,n} \mapsto \mathcal{Y}_i$ ,  $\forall i = 0, 1, \dots, n$ , is called causal if the map  $x^n \mapsto f_i(x^n)$  is measurable  $\forall i \in \mathbb{N}$  and

$$f_i(x^n) = f_i(\hat{x}^n) \quad \text{whenever} \quad x^i = \hat{x}^i, \quad \forall n \geq i, \quad n \in \mathbb{N}.$$

Since the class of randomized reproduction coders embeds deterministic causal reproduction coders, then probabilistically, a reproduction coder is causal if and only if the following MC holds  $X_{i+1}^\infty \leftrightarrow X^i \leftrightarrow Y_i$ ,  $\forall i \in \mathbb{N}$ . By (I.6), nonanticipatory  $\epsilon$ -entropy,  $R_{0,n}^\epsilon(D)$ , imposes a probabilistic causality (nonanticipative) constraint on the optimal reproduction distribution.

Next, we show that  $R_{0,n}^\epsilon(D) = R_{0,n}^{na}(D)$  and  $R^\epsilon(D) = R^{na}(D)$ , by invoking the following equivalent statements of MCs.

**Lemma III.2.** (Equivalent nonanticipative statements)

The following statements are equivalent.

- MC1:**  $P_{Y^n|X^n}(dy^n|x^n) = \overrightarrow{Q}_{Y^n|X^n}(dy^n|x^n) = \otimes_{i=0}^n P_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i)$ ,  $\forall n \in \mathbb{N}$ ;
- MC2:**  $Y_i \leftrightarrow (X^i, Y^{i-1}) \leftrightarrow (X_{i+1}, X_{i+2}, \dots, X_n)$  forms a MC, for each  $i = 0, 1, \dots, n-1$ ,  $\forall n \in \mathbb{N}$ ;
- MC3:**  $Y^i \leftrightarrow X^i \leftrightarrow X_{i+1}$  forms a MC, for each  $i = 0, 1, \dots, n-1$ ,  $\forall n \in \mathbb{N}$ ;
- MC4:**  $X_{i+1}^n \leftrightarrow X^i \leftrightarrow Y^i$  forms a MC, for each  $i = 0, 1, \dots, n-1$ ,  $\forall n \in \mathbb{N}$ .

*Proof:* See Appendix A. □

We note that **MC3** of Lemma III.2 is precisely Granger's definition of temporal causality [34], which is used in econometrics to unravel complex relations between macroeconomic variables from time series observations. It is also applied in bioengineering [34], [35], and more recently in neuroimaging to infer that  $\{Y_n : n \in \mathbb{N}\}$  does not cause  $\{X_n : n \in \mathbb{N}\}$ . Note that [36] refers to **MC4** as the "weak union" property of conditional independence.

In the next theorem, we utilize Lemma III.2, and more specifically, the fact that **MC4** is equivalent to **MC2** and **MC1**, to show that the extremum of the nonanticipatory  $\epsilon$ -entropy,  $R_{0,n}^\epsilon(D)$ , defined by (I.7), is equivalent to the extremum of nonanticipative RDF,  $R_{0,n}^{na}(D)$ , defined by (II.7).

**Theorem III.3.** (Equivalence of  $R_{0,n}^{na}(D)$  and  $R_{0,n}^\epsilon(D)$ )

The nonanticipative RDF and nonanticipatory  $\epsilon$ -entropy are equivalent notions, i.e.,  $R_{0,n}^{na}(D) = R_{0,n}^\epsilon(D)$ ,  $\forall n \in \mathbb{N}$ .

*Proof:* The derivation follows directly from Lemma III.2. □

### B. Infinite Horizon

In this section, we recall the finite horizon existence result derived in [16, Theorem III.4] to investigate existence of the information nonanticipative RDF rate, and the validity of interchanging the limit and infimum operations in (I.9). One may also consider the two-sided definition of nonanticipative RDF by replacing  $R_{0,n}^{na}(D)$  by  $R_{n_1,n_2}^{na}(D)$ ,  $n_2 > n_1$ , in which case, the rate is defined by  $\lim_{n_2-n_1 \rightarrow \infty} \frac{1}{n_2-n_1+1} R_{n_1,n_2}^{na}(D)$ , provided the limit exists. However, the rate is defined if and only if the following limit is defined for some  $n_1$ :  $\lim_{n_2 \rightarrow \infty} \frac{1}{n_2-n_1} R_{n_1,n_2}^{na}(D)$ . Hence, without loss of generality, we let  $n_1 = 0$ . First, we prove the following inequality.

**Lemma III.4.**

*The following holds.*

$$R^{na}(D) \leq \bar{R}^{na}(D) \triangleq \inf_{\vec{Q}_{Y^\infty|X^\infty}(\cdot|x^\infty) \in \vec{\mathcal{Q}}_{0,\infty}(D)} \lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n}). \quad (\text{III.3})$$

*Proof:* If the infimum in (II.7) does not exist, there is nothing to prove. Hence, suppose this infimum exists. By definition we have

$$R_{0,n}^{na}(D) \leq \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n}), \quad \forall \vec{Q}_{Y^n|X^n}(\cdot|x^n) \in \vec{\mathcal{Q}}_{0,n}(D)$$

and hence, by taking the limit on both sides we obtain

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na}(D) \leq \lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n}), \quad \forall \vec{Q}_{Y^\infty|X^\infty}(\cdot|x^\infty) \in \vec{\mathcal{Q}}_{0,\infty}(D).$$

Taking the infimum over  $\vec{Q}_{Y^\infty|X^\infty}(\cdot|x^\infty) \in \vec{\mathcal{Q}}_{0,\infty}(D)$  we obtain (III.3). This shows that  $R^{na}(D) \leq \bar{R}^{na}(D)$ .  $\square$

Since by Theorem III.3,  $R_{0,n}^{na}(D) = R_{0,n}^\varepsilon(D)$ ,  $\forall n \in \mathbb{N}$ , and since existence (under certain assumptions) of solution to the information nonanticipative RDF (II.7) is shown in [16, Theorem III.4], all technical results derived in [19, Theorems 1-4] are directly applicable to  $R_{0,n}^{na}(D)$  and its rate,  $R^{na}(D)$ , without assuming finiteness of  $R_{0,n}^{na}(D)$  for some  $n$ , as in [19]. Next, we summarize these results in order to conclude that the limit  $\lim_{n \rightarrow \infty} \frac{1}{n+1} R^{na}(D)$  is finite.

The following theorem is a direct consequence of Lemma III.4, [16, Theorem III.4], and Gorbunov-Pinsker [19, Theorem 2].

**Theorem III.5. (Limits)**

*Suppose the conditions of [16, Theorem III.4] holds and the source is stationary as defined in [19, Theorem 2]. Then the following hold.*

$$R^{na}(D) = \lim_{n \rightarrow \infty} \inf_{\vec{Q}_{Y^n|X^n}(\cdot|x^n) \in \vec{\mathcal{Q}}_{0,n}(D)} \frac{1}{n+1} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n}) < \infty \quad (\text{III.4})$$

*e.g., the limit exists and it is finite. Moreover,*

$$R^{na}(D) = \bar{R}^{na}(D) \equiv \inf_{\vec{Q}_{Y^\infty|X^\infty}(\cdot|x^\infty) \in \vec{\mathcal{Q}}_{0,\infty}(D)} \lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n}). \quad (\text{III.5})$$

*Proof:* The derivation utilizes the existence result derived in [16, Theorem III.4], and the subadditivity of  $R_{0,n}^{na}(D)$ , that is,  $R_{0,n}^{na}(D) \leq R_{0,k}^{na}(D) + R_{k+1,n}^{na}(D)$ ,  $0 < k < n$ . Since by Theorem III.3 we have  $R_{0,n}^{na}(D) = R_{0,n}^\varepsilon(D)$ , and by [19, Lemma 1],  $R_{0,n}^\varepsilon(D)$  is subadditive, then  $R_{0,n}^{na}(D)$  is also subadditive. Moreover, under the conditions of [16, Theorem III.4] we know that  $R_{0,n}^{na}(D)$  is finite for any finite  $n \in \mathbb{N}$ . Utilizing this and the subadditivity of  $R_{0,n}^{na}(D)$  we deduce that the limit in (III.4) exist and it is finite. (III.5) follows from [19, Theorem 2], the stationarity of the source, and Theorem III.3, which states that  $R_{0,n}^{na}(D) = R_{0,n}^\varepsilon(D)$ .  $\square$

Often, in the derivation of classical RDF rate it is assumed that the process  $\{(X_n, Y_n) : n \in \mathbb{N}\}$  is jointly stationary. This is not very natural because one does not know a priori whether the reproduction process is stationary. The next theorem utilizes [16, Theorem III.4] and the main theorems in [19] to establish that the infimum of the information nonanticipative RDF rate is achieved by a stationary reproduction distribution, which implies the process  $\{(X_n, Y_n) : n \in \mathbb{N}\}$  is jointly stationary.

**Theorem III.6. (Stationarity of the optimal reproduction distribution)**

*Suppose the conditions of [16, Theorem III.4] hold, the source  $\{X_n : n \in \mathbb{N}\}$  is stationary and consistent, and the fidelity set is shift invariant as defined in [19].*

*Then, the infimum in (III.5) is achieved by some  $\vec{Q}_{Y^n|X^n}^*(\cdot|x^n) \in \vec{\mathcal{Q}}_{0,n}(D)$  such that the source-reproduction pair  $\{(X_n, Y_n) : n \in \mathbb{N}\}$  is jointly stationary.*

*Proof:* By [16, Theorem III.4] the infimum is achieved, and by Theorem III.3, we have  $R_{0,n}^{na}(D) = R_{0,n}^\varepsilon(D)$ . By invoking [19, Theorem 4] we establish the claim of stationarity of the joint process.  $\square$

Therefore, by utilizing [16, Theorem III.4] we have strengthened the results described in [19, Theorems 1-4], which are based on the assumption that  $R_{0,n}^\varepsilon(D)$  is finite for some finite  $n \in \mathbb{N}$  and the infimum is achieved.

#### IV. OPTIMAL REPRODUCTION OF NONANTICIPATIVE RDF, PROPERTIES, AND EXAMPLES

In this section, we use the closed form expression of the optimal stationary reproduction corresponding to  $R^{na}(D)$  [16, Section IV] to derive an alternative characterization of the solution to the nonanticipative RDF,  $R^{na}(D)$ , and to introduce certain of its properties. Then we apply them to compute  $R^{na}(D)$  for the two running examples, the BSMS( $p$ ) and the general multidimensional partially observed stationary Gaussian-Markov source defined by (I.11).

The main assumption we impose is the following.

**Assumption IV.1.**

The  $(n+1)$ -fold convolution of causal conditional distribution  $\vec{Q}_{Y^n|X^n}(\cdot|x^n) = \otimes_{i=0}^n Q_{Y_i|Y^{i-1}, X^i}(\cdot|y^{i-1}, x^i)$  which achieves the infimum of  $R_{0,n}^{na}(D)$ , is a convolution of stationary conditional distributions.

**A. Stationary optimal reproduction distribution**

Clearly, (II.7) is a constrained problem, which is convex due to the convexity of the fidelity set, and the convexity of  $\mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \cdot)$ , as a functional of  $\vec{Q}_{Y^n|X^n}(\cdot|\cdot) \in \vec{\mathcal{Q}}(\mathcal{Y}_{0,n}|\mathcal{X}_{0,n})$ , (see Theorem II.2). Therefore, we apply duality theory [37] to convert the constrained problem into an unconstrained problem using Lagrange multipliers, and then we verify the equivalence of the constrained and unconstrained problems. This procedure is done in [16, Theorem IV.3] hence, it is omitted; instead we state the main theorem.

**Theorem IV.2. (Optimal stationary reproduction distribution)**

Suppose Assumption IV.1 and the conditions of [16, Theorem IV.4] hold, and  $d_{0,n} = \sum_{i=0}^n \rho(T^i x^n, T^i y^n)$ ,  $T^i x^n \subset \{x_0, \dots, x_i\}$ ,  $T^i y^n \subset \{y_0, \dots, y_i\}$ ,  $i = 0, \dots, n$ .

The following hold.

- (1) The infimum is attained at  $\vec{Q}_{Y^n|X^n}^*(\cdot|x^n) \in \vec{\mathcal{Q}}_{0,n}(D)$  given by<sup>4</sup>

$$\vec{Q}_{Y^n|X^n}^*(\times_{i=0}^n B_i|x^n) = \int_{B_0} Q_{Y_0|X_0}^*(dy_0|x_0) \dots \int_{B_n} Q_{Y_n|Y^{n-1}, X^n}^*(dy_n|y^{n-1}, x^n) \quad (\text{IV.1})$$

where

$$Q_{Y_i|Y^{i-1}, X^i}^*(dy_i|y^{i-1}, x^i) = \frac{e^{s\rho(T^i x^n, T^i y^n)} P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1})}{\int_{\mathcal{Y}_i} e^{s\rho(T^i x^n, T^i y^n)} P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1})}, \quad i = 0, 1, \dots, n, \quad s \leq 0 \quad (\text{IV.2})$$

and  $P_{Y_i|Y^{i-1}}^*(\cdot|\cdot) \in \mathcal{Q}_i(\mathcal{Y}_i|\mathcal{Y}_{0,i-1})$ ,  $i = 0, 1, \dots, n$ .

- (2) The information nonanticipative RDF is given by

$$\begin{aligned} R_{0,n}^{na}(D) &= sD(n+1) - \sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,n-1}} \log \left( \int_{\mathcal{Y}_i} e^{s\rho(T^i x^n, T^i y^n)} P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1}) \right) \\ &\quad \times \vec{Q}_{Y^{i-1}|X^{i-1}}^*(dy^{i-1}|x^{i-1}) \otimes P_{X^i}(dx^i). \end{aligned} \quad (\text{IV.3})$$

Moreover, if  $R_{0,n}^{na}(D) > 0$  then  $s < 0$ , and

$$\frac{1}{n+1} \sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i}} \rho(T^i x^n, T^i y^n) \vec{Q}_{Y^i|X^i}^*(dy^i|x^i) \otimes P_{X^i}(dx^i) = D. \quad (\text{IV.4})$$

*Proof:* The proof is described in [16, Theorem IV.4]. □

The RHS term of (IV.2) determines, for each  $i = 0, 1, \dots$ , the dependence of the reproduction distribution  $Q_{Y_i|Y^{i-1}, X^i}^*(\cdot|y^{i-1}, x^i)$  on the past reproduction  $y^{i-1}$  and the past and present source symbols  $x^i$ . Below we list a few observations regarding the information structure of the optimal stationary reproduction conditional distribution corresponding to  $R_{0,n}^{na}(D)$ . We shall use some of these in specific examples.

**Remark IV.3. (Information structures of the optimal stationary reproduction distribution)**

- (1) If  $\{X_n : n \in \mathbb{N}\}$  is stationary Gaussian and  $\rho(T^i x^n, T^i y^n) = \|x_i - y_i\|^2$ , a quadratic function of  $(x^n, y^n)$ , then for each  $(y^{i-1}, x^i) \in \mathcal{Y}_{0,i-1} \times \mathcal{X}_{0,i}$ ,  $Q_{Y_i|Y^{i-1}, X^i}^*(\cdot|y^{i-1}, x^i)$  is Gaussian. This follows from the fact that the exponent in the RHS of (IV.2) is quadratic in  $(x_i, y_i) \in \mathcal{X}_i \times \mathcal{Y}_i$ , and thus by assuming  $P_{Y_i|Y^{i-1}}^*(\cdot|y^{i-1})$  is conditionally Gaussian then the RHS of (IV.2) will be of exponential quadratic form in  $(x_i, y^i)$ . Hence, this RHS can be matched by a conditional Gaussian distribution for  $Q_{Y_i|Y^{i-1}, X^i}^*(\cdot|y^{i-1}, x^i)$ . The procedure is standard and involves completion of squares.
- (2) If the distortion function is  $\rho(T^i x^n, T^i y^n) = \rho(x_i, y^i)$  then

$$Q_{Y_i|Y^{i-1}, X^i}^*(\cdot|y^{i-1}, x^i) = Q_{Y_i|Y^{i-1}, X^i}^*(\cdot|y^{i-1}, x_i) - a.a. (x^i, y^{i-1}), \quad i = 0, 1, \dots$$

<sup>4</sup>Due to stationarity assumption  $P_{Y_i|Y^{i-1}}^*(\cdot|y^{i-1}) = P^*(\cdot|y^{i-1})$  and  $Q_{Y_i|Y^{i-1}, X^i}^*(\cdot|y^{i-1}, x^i) = Q^*(\cdot|y^{i-1}, x^i)$ .

That is, the reproduction conditional distribution is Markov in  $\{X_n : n \in \mathbb{N}\}$ . However, even if we further restrict the distortion function to single letter  $\rho(x_i, y_i)$ , we cannot deduce how far into the past  $Q_{Y_i|Y^{i-1}, X^i}^*(\cdot|y^{i-1}, x_i)$  depends on the reproduction symbols  $y^{i-1}$ . If the distortion function is of the form  $\rho(x_i, x_{i-1}, y^i)$  then

$$Q_{Y_i|Y^{i-1}, X^i}^*(\cdot; y^{i-1}, x^i) = Q_{Y_i|Y^{i-1}, X_i, X_{i-1}}^*(\cdot|y^{i-1}, x_i, x_{i-1}) - a.a.(x^i, y^{i-1}), \quad i = 0, 1, \dots$$

Despite the above observations, for specific sources with memory, we need additional properties of the solution of the nonanticipative RDF to determine the dependence of the optimal reproduction distribution on past reproduction symbols. The main properties are introduced below.

**Lemma IV.4.** (Convexity and monotonicity of  $R^{na}(D)$ )

$R_{0,n}^{na}(D)$  is a convex, nonincreasing function of  $D \in [0, \infty)$ .

*Proof:* By Theorem II.1 the set  $\vec{\mathcal{Q}}_{0,n}(D)$  is convex, and by Theorem II.2,  $\mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{\mathcal{Q}}_{Y^n|X^n})$  is a convex functional of  $\vec{\mathcal{Q}}_{Y^n|X^n}(\cdot|x^n) \in \vec{\mathcal{Q}}(\mathcal{Y}_{0,n}|\mathcal{X}_{0,n})$  for a fixed  $P_{X^n}(dx^n) \in \mathcal{M}(\mathcal{X}_{0,n})$ . Hence, the result follows.  $\square$   
In the next lemma we identify minimum value of  $D$  called  $D_{max}$ , beyond which  $R_{0,n}^{na}(D) = 0$ .

**Lemma IV.5.** ( $D_{max}$ )

$R_{0,n}^{na}(D) > 0$  for all  $D < D_{max}$  and  $R_{0,n}^{na}(D) = 0$  for all  $D \geq D_{max}$ , where

$$D_{max} \triangleq \min_{y^n \in \mathcal{Y}_{0,n}} \frac{1}{n+1} \sum_{i=0}^n \int_{\mathcal{X}_{0,i}} \rho(T^i x^n, T^i y^n) P_{X^i}(dx^i)$$

provided the minimum exists.

*Proof:* The derivation is similar to the one for the classical RDF, hence it is omitted.  $\square$

In the next lemma we give an alternative equivalent characterization of the optimal reproduction conditional distribution.

**Lemma IV.6.** (Equivalent characterization of optimal stationary reproduction)

The solution to the minimization problem of nonanticipative RDF defined by (IV.3) is such that

$$\int_{\mathcal{X}_{0,i}} e^{s\rho(T^i x^n, T^i y^n)} \lambda_i(x^i, y^{i-1}) P_{X^i|Y^{i-1}}^*(dx^i|y^{i-1}) = 1, \quad P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1}) - a.s., \quad \forall i \in \mathbb{N} \quad (IV.5)$$

where

$$\lambda_i(x^i, y^{i-1}) = \left( \int_{\mathcal{Y}_i} e^{s\rho(T^i x^n, T^i y^n)} P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1}) \right)^{-1}, \quad \forall i \in \mathbb{N} \quad (IV.6)$$

and  $P_{X^i|Y^{i-1}}^*(\cdot|\cdot) \in \mathcal{Q}(\mathcal{X}_{0,i}|\mathcal{Y}_{0,i-1})$ .

*Proof:* See Appendix B.  $\square$

Lemma IV.6 characterizes the solution to the optimization problem described in Theorem IV.2 on a set of  $P_{Y_i|Y^{i-1}}^*$ -measure 1. It can be shown, by utilizing measure theoretic arguments, that a necessary and sufficient condition for existence of a solution given in Theorem IV.2 is the condition

$$\int_{\mathcal{X}_{0,i}} e^{s\rho(T^i x^n, T^i y^n)} \lambda_i(x^i, y^{i-1}) P_{X^i|Y^{i-1}}^*(dx^i|y^{i-1}) \leq 1, \quad \forall y^i \in \mathcal{Y}_{0,i}, \quad i \in \mathbb{N}. \quad (IV.7)$$

Next, we utilize Lemma IV.6 to characterize the solution to the information nonanticipative RDF, as a maximization over a certain class of functions. By using this property, we can derive a lower bound on  $R_{0,n}^{na}(D)$ , which is analogous to Shannon's lower bound. In fact, we use this bound to determine the dependence of the optimal (stationary) reproduction distribution discussed in Remark IV.3 on past reproduction symbols, and to derive the information nonanticipative RDF of the multidimensional Gaussian-Markov process given by (I.11).

**Theorem IV.7.** (Alternative characterization of solution of the information nonanticipative RDF)

An alternative expression of the information nonanticipative RDF,  $R_{0,n}^{na}(D)$  is

$$R_{0,n}^{na}(D) = \sup_{s \leq 0} \sup_{\lambda \in \Psi_s} \left\{ sD(n+1) + \sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i-1}} \log \left( \lambda_i(x^i, y^{i-1}) \right) P_{X^{i-1}, Y^{i-1}}(dx^{i-1}, dy^{i-1}) \otimes P_{X_i|X^{i-1}}(dx_i|x^{i-1}) \right\} \quad (IV.8)$$

where

$$\Psi_s \triangleq \left\{ \lambda \triangleq \{ \lambda_i(x^i, y^{i-1}) \geq 0 : i = 0, 1, \dots, n \} : \int_{\mathcal{X}_{0,i}} e^{s\rho(T^i x^n, T^i y^n)} \lambda_i(x^i, y^{i-1}) P_{X^i|Y^{i-1}}(dx^i|y^{i-1}) \leq 1, \quad \forall y^i \in \mathcal{Y}_{0,i}, \quad i = 0, 1, \dots, n \right\}. \quad (IV.9)$$

Moreover, for each  $s \leq 0$ , a necessary and sufficient condition for  $\lambda_i(\cdot, \cdot)$  to achieve the supremum in (IV.8) is the existence of  $P_{Y_i|Y^{i-1}}^*(\cdot|\cdot)$  related to  $\lambda_i(\cdot, \cdot)$  via (IV.6) such that (IV.5) holds with equality a.a.  $y_i \in \mathcal{Y}_i$ , where  $P_{Y_i|Y^{i-1}}(dy_i|y^{i-1}) > 0$ ,  $i = 0, 1, \dots, n$ .

*Proof:* See Appendix C.  $\square$

The important point to be made regarding (IV.8) is that by removing the supremum over  $\lambda \in \Psi_s$  then a lower bound is obtained. For a given source and distortion function, this lower bound is then shown to be achievable by an optimal reproduction conditional distribution, which depends on finite past reproduction symbols.

#### B. Example 1: Binary Symmetric Markov Source (BSMS(p))

Consider a BSMS(p), with stationary transition probabilities  $\{P_{X_i|X_{i-1}}(x_i|x_{i-1}) : (x_i, x_{i-1}) \in \{0, 1\} \times \{0, 1\}\}$  given by  $P_{X_i|X_{i-1}}(0|0) = P_{X_i|X_{i-1}}(1|1) = 1 - p$ , and  $P_{X_i|X_{i-1}}(1|0) = P_{X_i|X_{i-1}}(0|1) = p$ ,  $i \in \{0, 1, \dots\}$ . We consider a single letter Hamming distortion criterion,  $\rho(x_i, y_i) = 0$  if  $x_i = y_i$  and  $\rho(x_i, y_i) = 1$  if  $x_i \neq y_i$ . The  $R^{na}(D)$  of the BSMS(p) is given in the next theorem.

**Theorem IV.8.** *The nonanticipative RDF,  $R^{na}(D)$ , for the BSMS(p) with single letter Hamming distortion function is*

$$R^{na}(D) = \begin{cases} H(m) - H(D) & \text{if } D \leq \frac{1}{2}, m = 1 - p - D + 2pD \\ 0 & \text{otherwise} \end{cases}$$

*the optimal (stationary) reproduction distribution is*

$$Q_{Y_i|Y_{i-1}, X_i}^*(y_i|y_{i-1}, x_i) = \begin{bmatrix} \alpha & 1 - \beta & \beta & 1 - \alpha \\ 1 - \alpha & \beta & 1 - \beta & \alpha \end{bmatrix} \quad (\text{IV.10})$$

where

$$\alpha = \frac{(1-p)(1-D)}{1-p-D+2pD}, \quad \beta = \frac{p(1-D)}{p+D-2pD}. \quad (\text{IV.11})$$

and  $\{Y_i : i = 0, 1, \dots\}$  is a first-order Markov with the same transition probability as that of the source.

Next, we outline the major steps of the proof; for a more detailed version see [38, Theorem 2.34].

*Proof:* The steady state distribution of the source is given by  $P_X(X_i = 0) \equiv P_X(0) = P_X(X_i = 1) \equiv P_X(1) = 0.5$ . By Theorem IV.8 (see Remark IV.3), the stationary reproduction distribution is Markov with respect to the source, given by

$$Q_{Y_i|Y^{i-1}, X_i}^*(y_i|y^{i-1}, x_i) = Q_{Y_i|Y^{i-1}, X_i}^*(y_i|y^{i-1}, x_i) = \frac{e^{s\rho(x_i, y_i)} P_{Y_i|Y^{i-1}}^*(y_i|y^{i-1})}{\sum_{y_i \in \{0, 1\}} e^{s\rho(x_i, y_i)} P_{Y_i|Y^{i-1}}^*(y_i|y^{i-1})}, \quad i = 0, 1, \dots \quad (\text{IV.12})$$

The distribution of the reproduction symbols,  $P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1})$ , is calculated by reconditioning on  $X_i$ . Then, it is substituted into the RHS OF (IV.12) to deduce that both  $P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1})$  and  $Q_{Y_i|Y^{i-1}, X_i}^*(y_i|y^{i-1}, x_i)$  are conditional independent of  $Y^{i-2}$ . Solving the system of resulting equations we calculate  $Q_{Y_i|Y^{i-1}, X_i}^*(y_i|y_{i-1}, x_i)$  as a function of the Lagrange multiplier “s”. The Lagrange multiplier “s” is found from the fidelity constraint and is substituted into the equation of the optimal reproduction distribution to obtain (IV.10). It can be verified that  $\{Y_i : i = 0, 1, \dots\}$  is first-order Markov with the same transition probability as that of the source  $\{X_i : i = 0, 1, \dots\}$ . Finally, the information nonanticipative RDF is computed using the expression

$$R^{na}(D) = \sum_{x_i, y_i, y_{i-1}} P_{X_i, Y_i, Y_{i-1}}^*(x_i, y_i, y_{i-1}) \log \left( \frac{Q_{Y_i|Y_{i-1}, X_i}^*(y_i|y_{i-1}, x_i)}{P_{Y_i|Y_{i-1}}^*(y_i|y_{i-1})} \right). \quad (\text{IV.13})$$

$\square$

The graph of  $R^{na}(D)$  is illustrated in Fig. IV.6; it shows that, as  $p$  tends to  $\frac{1}{2}$ , and the source becomes less correlated then  $R^{na}(D)$  increases. Note that for  $p = \frac{1}{2}$ , then BSMS( $\frac{1}{2}$ ) is the IID Bernoulli source. Then,  $m = 1 - p - D + 2pD = 0.5$  and therefore  $R^{na}(D) = 1 - H(D)$ ,  $D < \frac{1}{2}$ , which as expected is equal to the R(D) of an IID Bernoulli source with single letter distortion. At high resolution corresponding to  $D \rightarrow 0$ , then  $\lim_{D \rightarrow 0} R^{na}(D) \simeq H(p)$  which is the entropy rate of the BSMS(p).

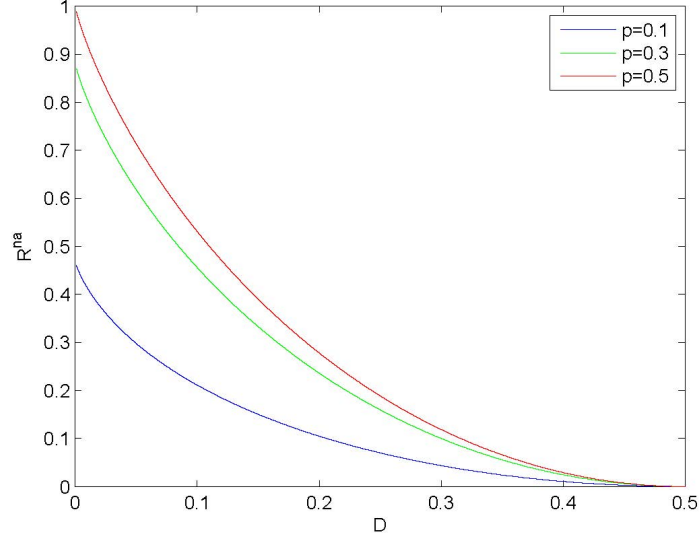


Fig. IV.6:  $R^{na}(D)$  for different values of parameter  $p$ .

### C. Example 2: Multidimensional Partially Observed Gaussian Process

Consider the discrete-time multidimensional partially observed linear Gauss-Markov source described by (I.11). The model in (I.11), is often encountered in applications where the process  $\{Z_t : t \in \mathbb{N}\}$  is not directly observed; instead, what is directly observed is the process  $\{X_t : t \in \mathbb{N}\}$  which is a noisy version of it. This is a realistic model for any sensor which collects information on the underlying process  $CZ_t$ , subject to additive Gaussian noise. Hence, in this application the objective is to compress the sensor data  $\{X_t : t = 0, 1, \dots, n\}$ . Next, we introduce certain assumptions concerning (I.11) and the distortion function, which are sufficient for existence of the limit,  $R^{na}(D) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na}(D)$ .

**(G1)**  $(C, A)$  is detectable and  $(A, \sqrt{BB^{tr}})$  is stabilizable, [28];

**(G2)** The state and observation noise  $\{(W_t, V_t) : t \in \mathbb{N}^n\}$  are Gaussian IID vectors  $W_t \in \mathbb{R}^k$ ,  $V_t \in \mathbb{R}^d$ , mutually independent with parameters  $N(0; I_{k \times k})$  and  $N(0; I_{d \times d})$ , independent of the Gaussian RV  $Z_0$ , with parameters  $N(0; \bar{\Sigma}_0)$ .

**(G3)** The distortion function is single letter defined by  $d_{0,n}(x^n, y^n) \triangleq \sum_{t=0}^n \|x_t - y_t\|_{\mathbb{R}^p}^2$ .

Special cases of model (I.11) are discussed in [39], [40]. Specifically, [39] computes  $R^{na}(D)$  for the scalar fully observed case corresponding to  $X_t = Z_t \in \mathbb{R}$  and [40] computes  $R^{na}(D)$  for the scalar partially observed case  $X_t \in \mathbb{R}$  via indirect methods.

Next, we invoke the characterization of the nonanticipative RDF given by Theorem IV.7 to derive the exact expression of  $R^{na}(D)$  for model (I.11). The corresponding optimal reproduction conditional distribution is constructed as shown in Fig. I.3. According to Theorem IV.2, the optimal stationary reproduction distribution is given by

$$\begin{aligned} Q_{Y_t|Y^{t-1}, X^t}^*(dy_t|y^{t-1}, x^t) &= \frac{e^{s\|y_t - x_t\|_{\mathbb{R}^p}^2} P_{Y_t|Y^{t-1}}^*(dy_t|y^{t-1})}{\int_{\mathcal{Y}_t} e^{s\|y_t - x_t\|_{\mathbb{R}^p}^2} P_{Y_t|Y^{t-1}}^*(dy_t|y^{t-1})}, \quad s \leq 0 \\ &\equiv Q_{Y_t|Y^{t-1}, X^t}^*(dy_t|y^{t-1}, x_t) - a.a. (y^{t-1}, x_t). \end{aligned} \quad (\text{IV.14})$$

Hence, from (IV.14), it follows that the optimal reproduction is Markov with respect to the process  $\{X_t : t \in \mathbb{N}\}$ . Moreover, since the exponential term  $\|y_t - x_t\|_{\mathbb{R}^p}^2$  in the RHS of (IV.14) is quadratic in  $(x_t, y_t)$ , and  $\{Z_t : i \in \mathbb{N}\}$  is Gaussian then  $\{(Z_t, X_t) : t \in \mathbb{N}\}$  are jointly Gaussian, and it follows that a Gaussian distribution  $Q_{Y_t|Y^{t-1}, X_t}(\cdot|y^{t-1}, x_t)$  (for a fixed realization of  $(y^{t-1}, x_t)$ ), and a Gaussian distribution  $P_{Y_t|Y^{t-1}}(\cdot|y^{t-1})$  can match the left and right side of (IV.14). Therefore, at any time  $t \in \mathbb{N}$ , the output  $Y_t$  of the optimal reconstruction channel depends on  $X_t$  and the previous outputs  $Y^{t-1}$ , and its conditional distribution is Gaussian. Hence, the channel connecting  $\{X_t : t \in \mathbb{N}\}$  to  $\{Y_t : t \in \mathbb{N}\}$  is realized by

$$Y_t = \bar{A}X_t + \bar{B}Y^{t-1} + V_t^c, \quad t \in \mathbb{N} \quad (\text{IV.15})$$

where  $\bar{A} \in \mathbb{R}^{p \times p}$ ,  $\bar{B} \in \mathbb{R}^{p \times tp}$ , and  $\{V_t^c : t \in \mathbb{N}\}$  is an independent sequence of Gaussian vectors  $N(0; Q_t)$ . Introduce the error estimate  $\{K_t : t \in \mathbb{N}\}$ , and its covariance  $\{\Lambda_t : t \in \mathbb{N}\}$ , defined by

$$K_t \triangleq X_t - \hat{X}_{t|t-1}, \quad \hat{X}_{t|t-1} \triangleq \mathbb{E}\{X_t | \sigma\{Y^{t-1}\}\}, \quad \Lambda_t \triangleq \mathbb{E}\{K_t K_t^{tr}\}, \quad t \in \mathbb{N} \quad (\text{IV.16})$$

where  $\sigma\{Y^{t-1}\}$  is the  $\sigma$ -algebra generated by the sequence  $\{Y^{t-1}\}$ . The covariance is diagonalized by introducing a unitary transformation  $\{E_t : t \in \mathbb{N}\}$  such that

$$E_t \Lambda_t E_t^{tr} = \text{diag}\{\lambda_{t,1}, \dots, \lambda_{t,p}\}, \quad \Gamma_t \triangleq E_t K_t, \quad t \in \mathbb{N}. \quad (\text{IV.17})$$

Note that although  $\{\Gamma_t : t \in \mathbb{N}\}$  has independent Gaussian components, each one of these components are correlated. Analogously, introduce to the process  $\{\tilde{K}_t : t \in \mathbb{N}\}$  defined by

$$\tilde{K}_t \triangleq Y_t - \hat{X}_{t|t-1}, \quad \tilde{\Gamma}_t = E_t \tilde{K}_t, \quad t \in \mathbb{N}. \quad (\text{IV.18})$$

Since  $d_{0,n}(X^n, Y^n) = d_{0,n}(K^n, \tilde{K}^n) = \sum_{t=0}^n \|\tilde{K}_t - K_t\|_{\mathbb{R}^p}^2 = \sum_{t=0}^n \|\tilde{\Gamma}_t - \Gamma_t\|_{\mathbb{R}^p}^2 = d_{0,n}(\Gamma^n, \tilde{\Gamma}^n)$  the square error fidelity criterion  $d_{0,n}(\cdot, \cdot)$  is not affected by the above preprocessing and post processing of  $\{(X_t, Y_t) : t \in \mathbb{N}\}$ . Moreover, using basic properties of conditional entropy, if necessary, we can show the following expressions are equivalent.

$$\begin{aligned} R^{na}(D) &= \lim_{n \rightarrow \infty} R_{0,n}^{na, K^n, \tilde{K}^n}(D) \triangleq \lim_{n \rightarrow \infty} \inf_{\tilde{P}_{\tilde{K}^n | K^n} : \mathbb{E}\{d_{0,n}(K^n, \tilde{K}^n) \leq D\}} \frac{1}{n+1} \sum_{t=0}^n I(K_t; \tilde{K}_t | \tilde{K}^{t-1}) \\ &= \lim_{n \rightarrow \infty} R_{0,n}^{na, \Gamma^n, \tilde{\Gamma}^n}(D) \triangleq \lim_{n \rightarrow \infty} \inf_{\tilde{P}_{\tilde{\Gamma}^n | \Gamma^n} : \mathbb{E}\{d_{0,n}(\Gamma^n, \tilde{\Gamma}^n) \leq D\}} \frac{1}{n+1} \sum_{t=0}^n I(\Gamma_t; \tilde{\Gamma}_t | \tilde{\Gamma}^{t-1}). \end{aligned} \quad (\text{IV.19})$$

Next, we give the expression of  $R^{na}(D)$  by using the specific realization of (IV.15) shown in Fig. I.3, where  $\{\mathcal{A}_\infty, \mathcal{B}_\infty : t = 0, 1, \dots\}$  are to be determined.

**Theorem IV.9.** ( $R^{na}(D)$  of multidimensional stationary partially observed Gaussian source)

Under Assumptions (G1)-(G3), the information nonanticipative RDF rate for the multidimensional stationary partially observed Gaussian source (I.11) is given by

$$R^{na}(D) = \frac{1}{2} \sum_{i=1}^p \log \left( \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} \right) \quad (\text{IV.20})$$

where  $\text{diag}\{\lambda_{\infty,1}, \dots, \lambda_{\infty,p}\} = \lim_{t \rightarrow \infty} E_t \Lambda_t E_t^{tr} = E_\infty \Lambda_\infty E_\infty^{tr}$ ,

$$\begin{aligned} \Lambda_\infty &= \lim_{t \rightarrow \infty} \mathbb{E} \left\{ \left( C(Z_t - \mathbb{E}\{Z_t | \sigma\{Y^{t-1}\}\}) + NV_t \right) \left( C(Z_t - \mathbb{E}\{Z_t | \sigma\{Y^{t-1}\}\}) + NV_t \right)^{tr} \right\} \\ &= C \lim_{t \rightarrow \infty} \Sigma_t C^{tr} + NN^{tr} = C \Sigma_\infty C^{tr} + NN^{tr} \end{aligned} \quad (\text{IV.21})$$

$$\delta_{\infty,i} \triangleq \begin{cases} \xi_\infty & \text{if } \xi_\infty \leq \lambda_{\infty,i} \\ \lambda_{\infty,i} & \text{if } \xi_\infty > \lambda_{\infty,i} \end{cases}, \quad i = 2, \dots, p \quad (\text{IV.22})$$

and  $\xi_\infty$  is chosen such that  $\sum_{i=1}^p \delta_{\infty,i} = D$ . Moreover,  $\Sigma_\infty$  is the steady state covariance of the error  $Z_t - \mathbb{E}\{Z_t | Y^{t-1}\} \sim N(0, \Sigma_\infty)$ ,  $\hat{Z}_{t|t-1} \triangleq \mathbb{E}\{Z_t | Y^{t-1}\}$ , of the Kalman filter given by

$$\begin{aligned} \hat{Z}_{t+1|t} &= A \hat{Z}_{t|t-1} \\ &\quad + A \Sigma_\infty (E_\infty^{tr} H_\infty E_\infty C)^{tr} M_\infty^{-1} (Y_t - C \hat{Z}_{t|t-1}), \quad \hat{Z}_{0|-1} = \mathbb{E}\{Z_0 | Y^{-1}\}, \quad Z_0 - \hat{Z}_{0|-1} \sim N(0, \Sigma_\infty), \quad t \in \mathbb{N} \end{aligned} \quad (\text{IV.23})$$

$$\Sigma_\infty = A \Sigma_\infty A^{tr} - A \Sigma_\infty (E_\infty^{tr} H_\infty E_\infty C)^{tr} M_\infty^{-1} (E_\infty^{tr} H_\infty E_\infty C) \Sigma_\infty A^{tr} + B B_\infty^{tr} \quad (\text{IV.24})$$

$$M_\infty = E_\infty^{tr} H_\infty E_\infty C \Sigma_\infty (E_\infty^{tr} H_\infty E_\infty C)^{tr} + E_\infty^{tr} H_\infty E_\infty N N^{tr} (E_\infty^{tr} H_\infty E_\infty)^{tr} + E_\infty^{tr} \mathcal{B}_\infty Q \mathcal{B}_\infty^{tr} E_\infty \quad (\text{IV.25})$$

and

$$H_\infty = \lim_{t \rightarrow \infty} H_t = \text{diag}\{\eta_{\infty,1}, \dots, \eta_{\infty,p}\}, \quad H_t \triangleq \text{diag}\{\eta_{t,1}, \dots, \eta_{t,p}\}, \quad \eta_{t,i} = 1 - \frac{\delta_{t,i}}{\lambda_{t,i}}, \quad i = 1, \dots, p, \quad t \in \mathbb{N} \quad (\text{IV.26})$$

$$\mathcal{B}_\infty = \lim_{t \rightarrow \infty} \mathcal{B}_t = \sqrt{H_\infty \Delta_\infty Q^{-1}}, \quad \mathcal{B}_t \triangleq \sqrt{H_t \Delta_t Q^{-1}}, \quad t \in \mathbb{N} \quad (\text{IV.27})$$

$$\Delta_\infty = \lim_{t \rightarrow \infty} \Delta_t = \text{diag}\{\delta_{\infty,1}, \dots, \delta_{\infty,p}\}, \quad \Delta_t = \text{diag}\{\delta_{t,1}, \dots, \delta_{t,p}\}, \quad t \in \mathbb{N} \quad (\text{IV.28})$$

$$Q = \lim_{t \rightarrow \infty} Q_t = \text{diag}\{q_{\infty,1}, \dots, q_{\infty,p}\}, \quad Q_t = \text{diag}\{q_{t,1}, \dots, q_{t,p}\}, \quad t \in \mathbb{N}. \quad (\text{IV.29})$$

Moreover,

$$Y_t = E_\infty^{tr} H_\infty E_\infty C (Z_t - \hat{Z}_{t|t-1}) + C \hat{Z}_{t|t-1} + E_\infty^{tr} H_\infty E_\infty N V_t + E_\infty^{tr} \mathcal{B}_\infty V_t^c, \quad t \in \mathbb{N}. \quad (\text{IV.30})$$

*Proof:* See Appendix D. □

For scalar stationary Gaussian sources with memory,  $R^{na}(D)$  given in Theorem IV.9 simplifies considerably. We illustrate this via examples.

#### Degraded cases.

- (1) Scalar Stationary Partially Observed Gaussian Markov Source: This corresponds to (I.11) by setting  $m = p = 1$ ,  $C = c$ ,  $N = \sigma_V$ ,  $A = \alpha$ ,  $B = \sigma_W$ , i.e.,  $\sigma_W W_t \sim N(0; \sigma_W^2)$  and  $\sigma_V V_t \sim N(0; \sigma_V^2)$  giving

$$\begin{cases} Z_{t+1} = \alpha Z_t + \sigma_W W_t, & t = 0, 1, \dots, \quad Z_0 \sim N(0; \sigma_W^2(1 - \alpha^2)^{-1}), \quad |\alpha| < 1, \\ X_t = cZ_t + \sigma_V V_t, & t = 0, 1, \dots \end{cases} \quad (IV.31)$$

Then  $\sigma_{Z_t}^2 \triangleq \text{Var}(Z_t) = \sigma_W^2(1 - \alpha^2)^{-1}$ ,  $\sigma_{X_t}^2 = c^2\sigma_{Z_t}^2 + \sigma_V^2$ . In this case, by Theorem IV.9 we have

$$\Lambda_\infty = \lambda_{\infty,1} = c^2\Sigma_\infty + \sigma_V^2, \quad \Delta_\infty = \delta_{\infty,1} = D, \quad \text{where } H_\infty = 1 - \frac{D}{c^2\Sigma_\infty + \sigma_V^2} \text{ and } E_\infty = 1.$$

Hence, using (IV.25), we obtain

$$M_\infty = c^2\Sigma_\infty H_\infty^2 + \sigma_V^2 H_\infty^2 + H_\infty D = H_\infty^2(c^2\Sigma_\infty + \sigma_V^2) + H_\infty D = (c^2\Sigma_\infty + \sigma_V^2)H_\infty. \quad (IV.32)$$

Also, from (IV.24), we obtain

$$\begin{aligned} \Sigma_\infty &= \alpha^2\Sigma_\infty - \alpha^2c^2\Sigma_\infty^2 H_\infty^2 M^{-1} + \sigma_W^2 \\ &\stackrel{(a)}{\implies} c^4\Sigma_\infty^3 + \Sigma_\infty^2(2c^2\sigma_V^2 - \alpha^2c^2\sigma_V^2 - \alpha^2c^2D - c^4\sigma_W^2) - \Sigma_\infty(\alpha^2\sigma_V^4 + 2c^2\sigma_V^2\sigma_W^2) - \sigma_V^4\sigma_W^2 = 0 \end{aligned} \quad (IV.33)$$

where (a) follows from (IV.32). It can be verified that the cubic equation (IV.33) admits a positive solution. From (IV.20) we obtain

$$R^{na}(D) = \frac{1}{2} \log \frac{\lambda_{\infty,1}}{\delta_{\infty,1}}, \quad \lambda_{\infty,1} = c^2\Sigma_\infty + \sigma_V^2 \geq D. \quad (IV.34)$$

- (2) Scalar Stationary Fully Observed Gaussian Markov Source: This corresponds to (IV.31) by setting  $c = 1$ ,  $\sigma_V = 0$  giving

$$X_{t+1} = \alpha X_t + \sigma_W W_t, \quad W_t \sim N(0; 1), \quad X_0 \sim N(0; \sigma_W^2(1 - \alpha^2)^{-1}), \quad |\alpha| < 1. \quad (IV.35)$$

Then  $\sigma_{X_t}^2 \triangleq \text{Var}(X_t) = \sigma_W^2(1 - \alpha^2)^{-1}$ . Since this is a special case of (1) for  $c = 1$ ,  $\sigma_V = 0$ , we obtain from (IV.32)

$$M_\infty = \Sigma_\infty H_\infty. \quad (IV.36)$$

Also, using (IV.33), we obtain

$$\lambda_{\infty,1} = \Sigma_\infty = \alpha^2 D + \sigma_W^2 \quad (IV.37)$$

Finally, by substituting (IV.37) in the expression of the nonanticipative RDF (IV.20) we obtain

$$R^{na}(D) = \frac{1}{2} \log \frac{\lambda_{\infty,1}}{\delta_{\infty,1}} = \frac{1}{2} \log \frac{\Sigma_\infty}{D} = \frac{1}{2} \log \left( \frac{\alpha^2 D + \sigma_W^2}{D} \right) = \frac{1}{2} \log \left( \alpha^2 + \frac{\sigma_W^2}{D} \right), \quad \Sigma_\infty \geq D, \quad |\alpha| < 1. \quad (IV.38)$$

This is precisely the expression derived in [15, Theorem 3] using power spectral densities.

- (3) IID Gaussian Source: This corresponds to (IV.35) by setting  $\alpha = 0$ ,  $\sigma_X = \sigma_W$ , which implies  $\{X_t : t = 0, \dots\}$  is  $N(0; \sigma_X^2)$ . By (IV.38), with  $\alpha = 0$ ,  $\sigma_X = \sigma_W$  then  $R^{na}(D) = R(D) = \frac{1}{2} \log \frac{\sigma_X^2}{D}$ ,  $\sigma_X^2 \geq D$ .

This is the well known RDF of the IID Gaussian source [3].

- (4) Vector Source with Independent Component versions of (1), (2). The vector versions of (IV.31) and (IV.35) with independent spacial components,  $\{Z_t^1, \dots, Z_t^m\}$ , and  $\{X_t^1, \dots, X_t^p\}$  is a straight forward extension of (IV.34) and (IV.38).

#### V. JSCC DESIGN BASED ON SYMBOL-BY-SYMBOL TRANSMISSION OF GAUSSIAN SOURCES WITH MEMORY

In this section, we use the solution of the nonanticipative RDF  $R^{na}(D)$  of the multidimensional stationary Gaussian source with memory in JSCC design using symbol-by-symbol transmission over an AWGN channel with or without feedback. We also illustrate that Theorem IV.9 and its corresponding realization scheme shown in Fig. I.3 gives as a special case the Schalkwijk-Kailath coding scheme.

First, we show that  $\{\tilde{K}_t : t \in \mathbb{N}\}$  is the innovation process of  $\{Y_t : t \in \mathbb{N}\}$ , and hence the two processes generate the same  $\sigma$ -algebras (they contain the same information).



**Lemma V.1.** (Equivalence of information generated by  $\{Y_t : t = 0, \dots\}$  and  $\{\tilde{K}_t : t = 0, \dots\}$ )  
The following hold.

$$\mathcal{F}_{0,t}^Y \triangleq \sigma\{Y_s : s = 0, 1, \dots, t\} = \mathcal{F}_{0,t}^{\tilde{K}} \triangleq \sigma\{\tilde{K}_s : s = 0, 1, \dots, t\}, \forall t \in \mathbb{N}.$$

that is,  $\mathcal{F}_{0,t}^Y \subseteq \mathcal{F}_{0,t}^{\tilde{K}}$  and  $\mathcal{F}_{0,t}^{\tilde{K}} \subseteq \mathcal{F}_{0,t}^Y, \forall t \in \mathbb{N}$ .

*Proof:* Since  $\tilde{K}_s = Y_s - \mathbb{E}\{X_s|Y^{s-1}\}$ ,  $0 \leq s \leq t$ , then  $\mathcal{F}_{0,t}^{\tilde{K}} \subseteq \mathcal{F}_{0,t}^Y, \forall t \in \mathbb{N}$ . Hence, we need to show that  $\mathcal{F}_{0,t}^Y \subseteq \mathcal{F}_{0,t}^{\tilde{K}}, \forall t \in \mathbb{N}$ . The innovation process of  $\{Y_t : t \in \mathbb{N}\}$  is by definition (see Fig. I.3, (IV.18), (D.1))

$$\begin{aligned} I_t &= Y_t - \mathbb{E}\{Y_t|Y^{t-1}\} \\ &= E_\infty^{tr} H_\infty E_\infty \left( X_t - \mathbb{E}\{X_t|Y^{t-1}\} \right) + E_\infty^{tr} \mathcal{B}_\infty V_t^c + \mathbb{E}\{X_t|Y^{t-1}\} - \mathbb{E}\{X_t|Y^{t-1}\} \\ &= E_\infty^{tr} H_\infty E_\infty \left( X_t - \mathbb{E}\{X_t|Y^{t-1}\} \right) + E_\infty^{tr} \mathcal{B}_\infty V_t^c = \tilde{K}_t. \end{aligned} \quad (\text{V.1})$$

Since the innovation process  $\{I_s : s = 0, 1, \dots, t\}$  and the optimal reproduction process  $\{Y_s : s = 0, 1, \dots, t\}$  generates the same  $\sigma$ -algebras, then  $\mathcal{F}_{0,t}^I \subseteq \mathcal{F}_{0,t}^Y, \mathcal{F}_{0,t}^Y \subseteq \mathcal{F}_{0,t}^I$ , i.e.,  $\mathcal{F}_{0,t}^Y = \mathcal{F}_{0,t}^I$ , and hence, by (V.1) we also obtain  $\mathcal{F}_{0,t}^Y \subseteq \mathcal{F}_{0,t}^{\tilde{K}}, \forall t \in \mathbb{N}$ .  $\square$   
We now observe the following consequence of Lemma V.1.

**Remark V.2.**

By Lemma V.1, all conditional expectations with respect to the process  $\{Y_t : t = 0, 1, \dots\}$  can be replaced by conditional expectations with respect to the independent process  $\{\tilde{K}_t : t = 0, 1, \dots\}$ . Hence, the process  $\{K_t : t = 0, 1, \dots\}$  can be written as  $K_t = X_t - \mathbb{E}\{X_t|\sigma\{Y^{t-1}\}\} = X_t - \mathbb{E}\{X_t|\sigma\{\tilde{K}^{t-1}\}\}$ , while its reconstruction is given by

$$\tilde{K}_t = E_\infty^{tr} H_\infty E_\infty \left( X_t - \mathbb{E}\{X_t|\tilde{K}^{t-1}\} \right) + E_\infty^{tr} \mathcal{B}_\infty V_t^c = E_\infty^{tr} H_\infty E_\infty K_t + E_\infty^{tr} \mathcal{B}_\infty V_t^c, \quad t = 0, 1, \dots \quad (\text{V.2})$$

Furthermore, by Lemma V.1,  $K_t$  and  $\tilde{K}_t$  are independent of  $Y_0, \dots, Y_{t-1}$ , and  $\tilde{K}_0, \dots, \tilde{K}_{t-1}$ ,  $t = 0, 1, \dots$ . This property is analogous to the JSCC of a scalar RV over a scalar additive Gaussian noise channel with feedback [23, Theorem 5.6.1].

The realization of Fig. I.3 illustrates a ‘‘Duality of a Source and a Channel’’ [41], that of the multidimensional stationary Gaussian source process and the multidimensional memoryless AWGN channel. To give an operational meaning to this duality, based on the realization of  $R^{na}(D)$  of Fig. I.3, we need to ensure that end-to-end average distortion is achieved. We do this by using existing results on capacity of memoryless AWGN channels.

Consider the memoryless AWGN channels which appear in Fig. I.3, defined by  $B_t = A_t + V_t^c$ ,  $t = 0, \dots, n$ , with  $\{V_t^c \triangleq \text{Vector}\{V_{t,1}^c, V_{t,2}^c, \dots, V_{t,p}^c\} : t = 0, \dots, n\}$ ,  $N(0; Q_t)$  (Gaussian) with  $Q_t \triangleq \text{Cov}(V_t^c) = \text{diag}\{q_{t,1}, q_{t,2}, \dots, q_{t,p}\}$ ,  $\{A_t \triangleq \text{Vector}\{A_{t,1}, A_{t,2}, \dots, A_{t,p}\} : t = 0, \dots, n\}$ ,  $P_t = \text{Cov}(A_t)$ ,  $t = 0, \dots, n$ , and power constraint  $\frac{1}{n+1} \sum_{t=0}^n \mathbb{E}\|A_t\|_{\mathbb{R}^p}^2 = \frac{1}{n+1} \sum_{t=0}^n \text{Trace}(P_t) \leq P$ . It is known that the capacity of such a channel with or without feedback subject to a power constraint is the same, and it is given by  $C(P) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n+1} C_{0,n}(P)$ , where

$$C_{0,n}(P) = \sup_{P_{A^n} : \frac{1}{n+1} \mathbb{E}\{\sum_{t=0}^n \|A_t^2\|_{\mathbb{R}^p}\} \leq P} I(A^n; B^n) \quad (\text{V.3})$$

Further, it is known that the channel input distribution corresponding to the maximization of the right hand side of (V.3) satisfies  $P_{A_t|A^{t-1}} = P_{A_t}, t = 0, \dots, n$ , and  $\{A_t : t = 0, 1, \dots, n\}$  is Gaussian  $N(0; P_t)$ ,  $P_t \triangleq E\{A_t A_t^r\} = \text{Cov}(A_t)$ . Denote the eigenvalues of  $P_t$  by  $P_{t,1}, P_{t,2}, \dots, P_{t,p}$ , for  $t = 0, 1, \dots, n$ . Then (V.3) becomes

$$C_{0,n}(P) \equiv C_{0,n}(P_{t,1}^*, \dots, P_{t,p}^* : t = 0, \dots, n) = \frac{1}{\frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p P_{t,i} \leq P} \max \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=0}^p \log\left(1 + \frac{P_{t,i}}{q_{t,i}}\right). \quad (\text{V.4})$$

where  $\{P_{t,1}^*, \dots, P_{t,p}^*\}$  is the optimal allocation of power. The capacity is given by

$$\begin{aligned} C(P) &\equiv C(P_{\infty,1}^*, \dots, P_{\infty,p}^*) = \lim_{n \rightarrow \infty} C_{0,n}(P) \\ &= \lim_{n \rightarrow \infty} \frac{1}{\frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p P_{t,i} \leq P} \max \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=0}^p \log\left(1 + \frac{P_{t,i}}{q_{t,i}}\right) = \frac{1}{2} \sum_{i=0}^p \log\left(1 + \frac{P_{\infty,i}^*}{q_{\infty,i}}\right), \quad \sum_{i=1}^p P_{\infty,i}^* = P. \end{aligned} \quad (\text{V.5})$$

The solution can be found using standard techniques and corresponds to water-filling of parallel memoryless Gaussian channels. Next, we ensure  $R(D) = C(P)$ , end-to-end average distortion is satisfied, and the encoder operates at channel capacity. For

a given  $D \in [D_{min}, D_{max}] > 0$  there exists a power  $P \in [P_{min}, P_{max}]$  such that

$$\begin{aligned} R^{na}(D) &= \lim_{n \rightarrow \infty} R_{0,n}^{na}(D) = \lim_{n \rightarrow \infty} \frac{1}{2(n+1)} \sum_{t=0}^n \log \frac{|\Lambda_t|}{|\Delta_t|} \\ &= \frac{1}{2} \sum_{i=1}^p \log \left( \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} \right) = \frac{1}{2} \sum_{i=1}^p \log \left( 1 + \frac{P_{\infty,i}^*}{q_{\infty,i}} \right) = C(P) \Big|_{\frac{P_{\infty,i}^*}{q_{\infty,i}} = \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} - 1, i=1, \dots, p}. \end{aligned} \quad (V.6)$$

Below we give the JSCC design based on symbol-by-symbol transmission with and without feedback encoding.

**JSCC Design with Feedback.** The AWGN channel with feedback has the following implementation.

$$B_t = A_{\infty}^{fb}(X_t, B^{t-1}) + V_t^c = \mathcal{A}_{\infty} E_{\infty} \left( X_t - \mathbb{E}\{X_t | B^{t-1}\} \right) + V_t^c, \quad \mathcal{A}_{\infty} \triangleq \sqrt{Q \Delta_{\infty}^{-1} H_{\infty}}, \quad t \in \mathbb{N}, \quad i = 1, \dots, p. \quad (V.7)$$

where  $\mathcal{A}_{\infty}$  is chosen to ensure the power allocation is satisfied and to guarantee the encoder operates at  $C(P)$ . This shows that, for a given distortion level  $D \in [D_{min}, D_{max}] > 0$ , the realization shown in Fig. I.3 is optimal in the sense that the end-to-end nonanticipative RDF,  $R^{na}(D)$  is achieved (with the prescribed average distortion), the encoder (V.7) achieves the capacity of the channel, and (V.6) is satisfied. Thus,  $R^{na}(D)$  is achievable over the {encoder, channel, decoder} design shown in Fig. I.3.

**JSCC Design without Feedback.** The AWGN channel without feedback follows from (V.7) with  $\mathbb{E}\{X_t | B^{t-1}\}$  replaced by  $\mathbb{E}\{X_t | \sigma(\mathbb{R}^p)\} = EX_t = 0$ , that is,  $B_t = A_{\infty}^{fb}(X_t) + V_t^c = \mathcal{A}_{\infty} E_{\infty} X_t + V_t^c$ ,  $\mathcal{A}_{\infty} \triangleq \sqrt{Q \Delta_{\infty}^{-1} H_{\infty}}$ ,  $t \in \mathbb{N}$ ,  $i = 1, \dots, p$ .

Note that the JSCC design (shown in Fig. I.3) is not the only choice. One may consider JSCC design of the multidimensional Gaussian source with memory over other types of Gaussian channels.

Next, we show that *JSCC design of Multidimensional stationary Gaussian Sources with Memory* presented in Fig. I.3 includes as degraded special cases several scenarios, including previously known results.

#### Degraded cases.

**(FB1)** *JSCC design of transmitting a scalar Gaussian Markov Source over a memoryless AWGN channel with feedback.* Let  $X_t$  be a scalar Gaussian Markov source defined by (IV.35), where it is shown that  $\lambda_{\infty,1} = \alpha^2 D + \sigma_W^2$ ,  $R^{na}(D) = \frac{1}{2} \log(\alpha^2 + \frac{\sigma_W^2}{D})$  (see (IV.37), (IV.38)). The capacity of a memoryless AWGN channel with or without feedback is obtained from (V.5) by setting  $p = 1$  and  $q_{\infty,1} = \mathbb{E}\{|V_t^c|^2\} = Q = \sigma_{V_c}^2$  giving

$$C(P) = \frac{1}{2} \log \left( 1 + \frac{P}{\sigma_{V_c}^2} \right). \quad (V.8)$$

Suppose that the channel is used once per source symbol. For this realization, the smallest achievable distortion is

$$D_{min} = \frac{\sigma_W^2 \sigma_{V_c}^2}{(1 - \alpha^2) \sigma_{V_c}^2 + P}. \quad (V.9)$$

Using Theorem IV.9 and (V.7) we obtain

$$B_t = \sqrt{\frac{P}{\lambda_{\infty,1}}} K_t + V_t^c = \sqrt{\frac{P((1 - \alpha^2) \sigma_{V_c}^2 + P)}{\sigma_W^2 (\sigma_{V_c}^2 + P)}} K_t + V_t^c, \quad K_t = X_t - \mathbb{E}\{X_t | B^{t-1}\}. \quad (V.10)$$

From Remark V.2 (by setting  $p = 1$ ), the decoder expression (V.2) becomes

$$\tilde{K}_t = H_{\infty} K_t + \mathcal{B}_{\infty} V_t^c = \mathcal{B}_{\infty} (\mathcal{A}_{\infty} K_t + V_t^c) \stackrel{(a)}{=} \mathcal{B}_{\infty} B_t \quad (V.11)$$

where (a) follows from (V.7). By using the fact that  $q_{\infty,1} = \sigma_{V_c}^2$ ,  $\delta_{\infty,1} = D$  the scaling factor  $\mathcal{B}_{\infty}$  (IV.27) (for the scalar case  $p = 1$ ), which guarantees the minimum end-to-end distortion error is

$$\mathcal{B}_{\infty} = \sqrt{\frac{(1 - \frac{\delta_{\infty,1}}{\lambda_{\infty,1}}) \delta_{\infty,1}}{q_{\infty,1}}} = \sqrt{\frac{\sigma_W^2 P}{((1 - \alpha^2) \sigma_{V_c}^2 + P) (\sigma_{V_c}^2 + P)}}. \quad (V.12)$$

By substituting (V.12) into (V.11) we get

$$\tilde{K}_t = \sqrt{\frac{\sigma_W^2 P}{((1 - \alpha^2) \sigma_{V_c}^2 + P) (\sigma_{V_c}^2 + P)}} B_t. \quad (V.13)$$

Finally, the average end-to-end distortion is computed by evaluating the expectation

$$D = \mathbb{E}\{|X_t - Y_t|^2\} = \mathbb{E}\{|K_t - \tilde{K}_t|^2\} = \frac{\sigma_W^2 \sigma_{V_c}^4 + \sigma_W^2 \sigma_{V_c}^2 P}{((1 - \alpha^2) \sigma_{V_c}^2 + P)(P + \sigma_{V_c}^2)} = \frac{\sigma_W^2 \sigma_{V_c}^2}{(1 - \alpha^2) \sigma_{V_c}^2 + P} = D_{min}.$$

This JSCC design is the one illustrated in Fig. I.4.

**Special case: Realization Without Feedback.** When there is no feedback, all statements presented in **(FB1)** holds, with  $\mathbb{E}\{X_t|B^{t-1}\}$  replaced by  $\mathbb{E}\{X_t|\sigma\{\mathbb{R}\}\} = \mathbb{E}\{X_t\} = 0$  (i.e., only a priori information is used), and the encoder has the following structure.

$$B_t = \sqrt{\frac{P}{\lambda_{\infty,1}}} X_t + V_t^c, \quad \lambda_{\infty,1} = \text{Var}(X_t) = \sigma_W^2 (1 - |\alpha|^2)^{-1}, \quad |\alpha| < 1. \quad (\text{V.14})$$

Toward this, we inspect **(FB1)** without feedback.

**(NFB1)** *JSCC design of transmitting a scalar Gaussian Markov source over a memoryless AWGN channel without feedback.* Let  $\{X_t : t = 0, 1, \dots\}$  be the scalar stationary Gaussian Markov source defined by (IV.35). In contrary to the case **(FB1)**, here the encoder has the form of (V.14), where  $\lambda_{\infty,1} = \sigma_{X_t}^2 = \frac{\sigma_W^2}{1 - \alpha^2}$ , and  $R^{na}(D) = \frac{1}{2} \log \frac{\sigma_W^2}{(1 - \alpha^2)D}$ ,  $\frac{\sigma_W^2}{1 - \alpha^2} \geq D$ ,  $|\alpha| < 1$ . Consider the realization of  $R^{na}(D)$  over a memoryless AWGN channel without feedback whose channel capacity is defined by (V.8). Suppose that the channel is used once per source symbol. For this realization, the smallest achievable distortion is

$$D_{min} = \frac{\sigma_W^2 \sigma_{V_c}^2}{(1 - \alpha^2)(P + \sigma_{V_c}^2)}. \quad (\text{V.15})$$

Then (V.14) is expressed as

$$B_t = \sqrt{\frac{P}{\lambda_{\infty,1}}} K_t + V_t^c = \sqrt{\frac{(1 - \alpha^2)P}{\sigma_W^2}} X_t + V_t^c. \quad (\text{V.16})$$

By using the fact that  $q_{\infty,1} = \sigma_{V_c}^2$ ,  $\delta_{\infty,1} = D$  the scaling factor  $\mathcal{B}_{\infty}$  (IV.27) for the scalar case  $p = 1$ , which guarantees the minimum end-to-end distortion error is

$$\mathcal{B}_{\infty} = \sqrt{\frac{(1 - \frac{\delta_{\infty,1}}{\lambda_{\infty,1}}) \delta_{\infty,1}}{q_{\infty,1}}} = \sqrt{\frac{\lambda_{\infty,1}}{P}} \frac{P}{P + \sigma_{V_c}^2} = \sqrt{\frac{\sigma_W^2}{(1 - \alpha^2)P}} \frac{P}{P + \sigma_{V_c}^2}. \quad (\text{V.17})$$

By substituting (V.17) into (V.11) we get

$$\tilde{K}_t = \sqrt{\frac{\sigma_W^2}{(1 - \alpha^2)P}} \frac{P}{P + \sigma_{V_c}^2} B_t. \quad (\text{V.18})$$

Finally, the end-to-end distortion is computed by evaluating the expectation

$$D = \mathbb{E}\{|X_t - Y_t|^2\} = \mathbb{E}\{|K_t - \tilde{K}_t|^2\} = \frac{\sigma_W^2 \sigma_{V_c}^4}{(1 - \alpha^2)(P + \sigma_{V_c}^2)^2} + \frac{\sigma_W^2 \sigma_{V_c}^2 P}{(1 - \alpha^2)(P + \sigma_{V_c}^2)^2} = \frac{\sigma_W^2 \sigma_{V_c}^2}{(1 - \alpha^2)(P + \sigma_{V_c}^2)} = D_{min}.$$

This JSCC design is the one illustrated in Fig. I.5.

**(NFB2)** *JSCC design of transmitting a scalar IID Gaussian source over a memoryless AWGN channel without feedback.* By setting  $\alpha = 0$  and  $\sigma_X = \sigma_W$  in case **(NFB1)**, the encoder has the form of (V.14) where  $\lambda_{\infty,1} = \sigma_X^2$ .

This is the case discussed in [13] (see also [42, Example 2.2]).

Next, we show how to recover from the realization scheme depicted in Fig. I.3, the *Schalkwijk-Kailath coding scheme which achieves the feedback capacity of memoryless Gaussian channels* [18].

**Feedback Realization: The Schalkwijk-Kailath coding scheme.** Consider a scalar Gaussian RV  $X \sim N(0; \sigma_X^2)$ . By letting  $p = 1$  in (V.6)-(V.7), we have  $\mathcal{A}_{\infty} \equiv \sqrt{Q \Delta_{\infty}^{-1} H_{\infty}} = \sqrt{q_{\infty,1} \frac{1}{\delta_{\infty,1}} (1 - \frac{\delta_{\infty,1}}{\lambda_{\infty,1}})}$ ,  $\frac{P_{\infty,1}^*}{q_{\infty,1}} = \frac{\lambda_{\infty,1}}{\delta_{\infty,1}} - 1$  which implies  $\mathcal{A}_{\infty} = \sqrt{\frac{P_{\infty,1}^*}{\lambda_{\infty,1}}}$ ,

$\lambda_{\infty,1} = \text{Var}(X - \mathbb{E}\{X|B^{t-1}\}) = \text{Var}(K_t)$ ,  $K_t = X - \mathbb{E}\{X|B^{t-1}\}$  and  $B_t = \sqrt{\frac{P_{\infty,1}^*}{\lambda_{\infty,1}}} K_t + V_t^c, t \in \mathbb{N}$ .

Substituting into the encoder (V.7) the limiting values,  $P_{\infty,1}^* = P$  then

$$B_t = \sqrt{\frac{P}{\lambda_{\infty,1}}} (X - \mathbb{E}\{X|B^{t-1}\}) + V_t^c, \quad t = 0, 1, \dots \quad (\text{V.19})$$

This is the Schalkwijk-Kailath coding scheme [18] of a scalar Gaussian RV  $X$  in which an encoder is designed to achieve the capacity of memoryless AWGN channel with feedback.

**Remark V.3.**

The above JSCC designs based on symbol-by-symbol transmission of scalar sources over AWGN channels, can be generalized to vector sources over vector AWGN channels.

## VI. BOUNDS ON OPTA BY NONCAUSAL AND CAUSAL CODES

In this section, we show that the nonanticipative RDF is a lower bound on the OPTA by causal codes developed in [5]. Consider a causal source code [5] and define the average fidelity by

$$d^+(x, y) \triangleq \limsup_{k \rightarrow \infty} \frac{1}{n+1} \mathbb{E} \left\{ d_{0,n}(x^n, y^n) \right\}, \quad d_{0,n}(x^n, y^n) \triangleq \sum_{i=0}^n \rho(x_i, y_i).$$

Let  $l_n(x^\infty)$  denote the total number of bits received at the decoder at the time it reproduces the output sequence  $\{Y_n : n \in \mathbb{N}\}$ , when the source is  $\{X_n : n \in \mathbb{N}\}$ . In [5] the average rate of the encoder-decoder pairs using causal reproduction coders is measured by  $\limsup_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{E} \left\{ l_n(X^\infty) \right\}$ .

Moreover, given a source  $\{X_n : n = 0, 1, \dots\}$  the OPTA by causal codes subject to fidelity is given by [5]

$$r^{c,+}(D) \triangleq \inf_{\substack{y_i: y_i = f_i(x^i), \forall i \in \mathbb{N} \\ f_i \text{ is causal, } \forall i \in \mathbb{N}, d^+(x, y) \leq D}} \limsup_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{E} \left\{ l_n(X^\infty) \right\}. \quad (\text{VI.1})$$

Causal codes based on [5] are analyzed and further generalized in [43] for stationary ergodic sources, under a variety of side information available at the encoder and decoder. Although expression (VI.1) is very attractive, its computation for general sources is very difficult.

Next, we show that the OPTA by causal codes is bounded below by the expression of information nonanticipative RDF rate. Consider the joint distribution defined by  $P_{X^n}(dx^n)$ , and a reproduction distribution  $\vec{Q}_{Y^n|X^n}(\cdot|x^n)$ , so that the randomized coders are consistent with Definition III.1. Then, by data processing inequality we have the following bounds.

$$\mathbb{E} \left\{ l_n(X^\infty) \right\} \geq H(Y^n) \geq \sum_{i=0}^n \left\{ H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, X^i) \right\} \stackrel{(b)}{=} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n}) \quad (\text{VI.2})$$

where (b) follows from the fact that the joint distribution is defined by  $P_{X^n}(dx^n)$  and the conditional reproduction distribution  $\vec{Q}_{Y^n|X^n}(\cdot|x^n)$ . Therefore, by taking the infimum of the RHS of (VI.2) over  $\vec{Q}_{Y^n|X^n}(\cdot|x^n) \in \vec{\mathcal{Q}}_{0,n}(D)$  and its left side of reproduction codes as in (VI.1) we obtain

$$r_{0,n}^c(D) \triangleq \inf_{\substack{y_i: y_i = f_i(x^i), f_i \text{ is causal, } i=0,1,\dots,n \\ \frac{1}{n+1} \mathbb{E} \{ d_{0,n}(x^n, y^n) \} \leq D}} \frac{1}{n+1} \mathbb{E} \left\{ l_n(X^\infty) \right\} \geq \frac{1}{n+1} R_{0,n}^{na}(D) \stackrel{(c)}{\geq} \frac{1}{n+1} R_{0,n}(D)$$

where (c) follows from the fact that  $R_{0,n}^{na}(D) \geq R_{0,n}(D)$ .

In the previous bounds we can first take  $\limsup_{n \rightarrow \infty}$  and then the infimum giving

$$\begin{aligned} r^{c,+}(D) &\geq R^{na,+}(D) \triangleq \inf_{\substack{\vec{Q}_{Y^\infty|X^\infty}(\cdot|x^\infty) \\ \in \vec{\mathcal{Q}}_{0,\infty}(D)}} \limsup_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na}(D) \\ &\geq R^+(D) \triangleq \inf_{\substack{\vec{Q}_{Y^\infty|X^\infty}(\cdot|x^\infty) \\ \in \vec{\mathcal{Q}}_{0,\infty}(D)}} \limsup_{n \rightarrow \infty} \frac{1}{n+1} R(D). \end{aligned}$$

Therefore, the information nonanticipative RDF,  $R^{na}(D)$ , and rate  $R^{na,+}(D)$ , are lower bounds on  $r^{c,+}(D)$ , the OPTA by causal codes, and upper bounds to the classical RDF and rate  $R^+(D)$ .

*Bounds on the OPTA by causal codes for BSMS(p).*

The classical RDF for the BSMS(p) is only known for the specific distortion region  $0 \leq D \leq D_c$  [44], and is given by

$$R(D) = H(p) - H(D) \text{ if } D \leq D_c = \frac{1}{2} \left( 1 - \sqrt{1 - \left( \frac{p}{q} \right)^2} \right), \quad p \leq 0.5. \quad (\text{VI.3})$$

For the remainder of the distortion region  $D > D_c$  only upper and lower bounds on  $R(D)$  are known [27]. In fact, it is shown by Gray [44] that (VI.3) is also a lower bound for  $R(D)$  in the region  $D_c \leq D \leq \frac{1}{2}$  and is equivalent to the Shannon lower bound. Our expression of the nonanticipative RDF provides an upper bound on the classical RDF for all possible values of  $D$ ,  $0 \leq D \leq 0.5$ . Figure VI.7 shows the upper bound derived by Berger [27, equations 46, 47], which hold for  $D_c \leq D \leq \frac{1}{2}$ , Shannon's lower bound, and the upper bound based on  $R^{na}(D)$ . Shannon's lower bound is tight for  $D \leq D_c$  [44]. Moreover, since  $R^{na}(D)$  is nonincreasing and convex as a function of  $D$ , and nonincreasing for all values of  $p \in [0, 0.5]$

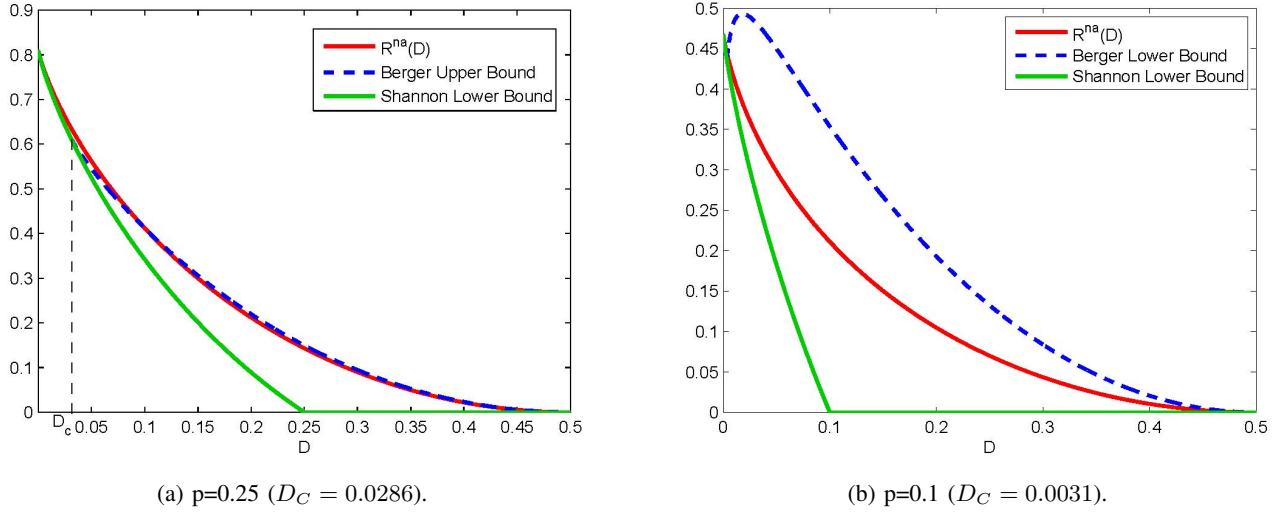


Fig. VI.7: Bounds for BSMS( $p$ ). Shannon's lower bound is tight for  $D < D_C$  [44].

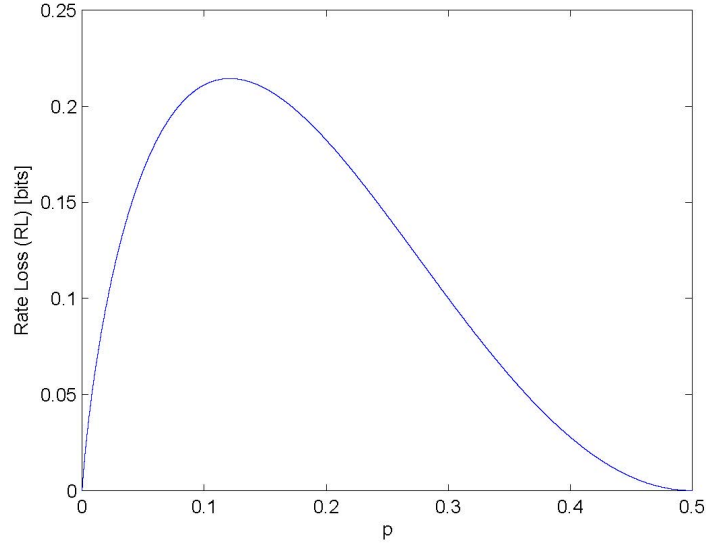


Fig. VI.8: Upper bound on the Rate Loss (RL) for the BSMS,  $p \in [0, 0.5]$ .

(these are easily shown), then the upper bound based on  $R^{na}(D)$  is convex, when compared to Berger's upper bound which is not necessarily convex and nonincreasing (as illustrated by the blue graph in Figure VI.7a). Finally, we use the bound  $R(D) \leq R^{na}(D) \leq r^{c,+}(D)$  to deduce that the RL of causal codes with respect to the OPTA by noncausal codes of the BSMS( $p$ ) cannot exceed

$$RL = R^{na}(D) - R(D) \leq \begin{cases} H(m) - H(p) & \text{if } 0 \leq D \leq p \\ H(m) - H(D) & \text{if } p < D \leq 0.5. \end{cases}$$

This bound on the RL is illustrated in Figure VI.8, and demonstrates the fluctuation of the RL for  $p \in [0, 0.5]$ . It is interesting to see that the maximum value of the RL is 0.2144 and corresponds to  $(p = 0.1012, D = 0.1012)$ . For high resolution ( $D \rightarrow 0$ ), the classical RDF and the nonanticipative RDF are equivalent and equal to  $H(p)$ .

*Bounds on multidimensional partially observed Gaussian source.*

For the multidimensional partially observable stationary Gaussian-Markov source given in (I.11), the RL of causal codes with respect to  $R(D)$  is at most  $R^{na}(D) - R(D)$  bits/sample, where  $R^{na}(D)$  is given in Theorem IV.9, while the expression for  $R(D)$  is found in [3]. On the other hand,  $R^{na}(D) - R(D)$  is the RL of zero-delay codes with respect to the values of  $R(D)$ . To facilitate the computation of RL of zero delay codes with respect to those of  $R(D)$ , one can work in frequency domain, by deriving the equivalent expression of  $R^{na}(D)$  using the solution given in Theorem IV.9 and Szegő formulae. For

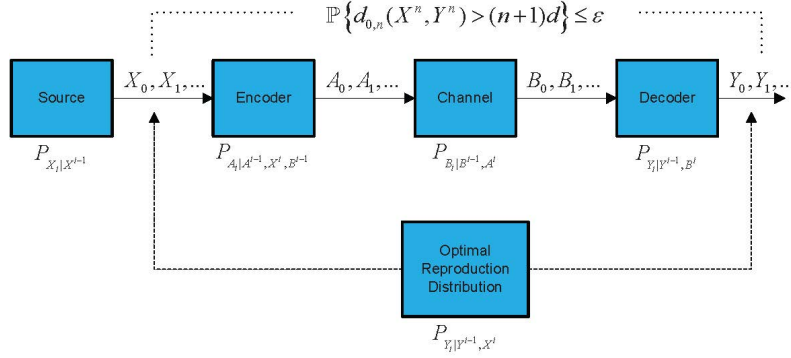


Fig. VII.9: JSCC design based on nonanticipative transmission for sources with memory.

scalar Gaussian stationary processes such an expression is given in [20], [21].

Next, we evaluate the RL of causal codes with respect  $R(D)$  by considering the first-order (scalar) Gaussian-Markov autoregressive source given by (IV.35). For this model we take  $\alpha = 1$  which is the model with a parametric expression for specific distortion region, discussed in [3, Example 6.3.2.1]. Specifically, for this model, the parametric expression of  $R(D)$  is  $R(D) = \frac{1}{2} \log \frac{\sigma_W^2}{D}$ ,  $0 \leq D \leq D_c = \frac{\sigma_W^2}{4}$ . For the nonanticipative RDF, this model can be explicitly computed by (IV.38) by setting  $\alpha = 1$  as follows.

$$R^{na}(D) = \frac{1}{2} \log \left( 1 + \frac{\sigma_W^2}{D} \right), \quad 0 \leq D \leq \infty.$$

Hence, the RL due to causal codes for the first order Gaussian-Markov autoregressive source given by (IV.35) for  $\alpha = 1$  cannot exceed

$$RL = R^{na}(D) - R(D) = \frac{1}{2} \log \left( 1 + \frac{D}{\sigma_W^2} \right), \quad 0 \leq D \leq D_c = \frac{\sigma_W^2}{4}.$$

Note for  $D > \frac{\sigma_W^2}{4}$ , no explicit closed form expression of  $R(D)$  for the first order Gaussian autoregressive source is found in the literature. Instead, only approximation solutions are given [3, Chapter 6].

## VII. CODING THEOREM FOR JSCC DESIGN USING NONANTICIPATIVE CODES

In this section, we describe a general constructive procedure for JSCC design based on nonanticipative transmission, for sources with memory and channels with memory with and without feedback, with respect to average end-to-end distortion or excess distortion probability, operating optimally, that is,

- (1) the end-to-end average or excess distortion probability is achieved;
- (2) the encoder achieves the channel capacity;
- (3)  $R^{na}(D) = C(P)$ , where  $C(P)$  is the capacity of the channel with power  $P$ .

The constructive procedure is a generalization of the JSCC design of multidimensional stationary Gaussian-Markov source transmitted over the vector memoryless AWGN channel presented in Section IV-C. For memoryless sources and memoryless channels a similar method is described in [14], and evaluated for *Example-IID-BSS* and *Example-IID-GS*.

### A. Nonanticipative JSCC Design

The elements of the JSCC design are illustrated in Fig. VII.9. We focus on JSCC design systems which are nonanticipative, that is, the encoder, channel, and decoder at each time instant  $i$  process samples causally, with memory on past symbols, and without anticipation with respect to symbols occurring at times  $j > i$ . Since by data processing inequality, information nonanticipative RDF is lower than the capacity of the channel, we impose a cost for transmitting symbols over the channel, which is a measurable function

$$c_{0,n}: \mathcal{A}_{0,n} \times \mathcal{B}_{0,n-1} \mapsto [0, \infty), \quad c_{0,n}(a^n, b^{n-1}) \triangleq \sum_{i=0}^n \gamma(T^i a^n, T^i b^{n-1}) \quad (\text{VII.1})$$

where  $T^i a^n \subset \{a_0, \dots, a^i\}$  and  $T^i b^{n-1} \subset \{b_0, \dots, b^{i-1}\}$ .

We use the following definition of a nonanticipative code.

**Definition VII.1.** (Nonanticipative code)

An  $(n, d, \epsilon, P)$  nonanticipative code is a tuple

$$\left( \mathcal{X}_{0,n}, \mathcal{A}_{0,n}, \mathcal{B}_{0,n}, \mathcal{Y}_{0,n}, P_{X^n}, \vec{P}_{A^n|B^{n-1}, X^n}, \vec{P}_{B^n|A^n}, \vec{P}_{Y^n|B^n}, d_{0,n}, c_{0,n} \right)$$

where  $\vec{P}_{A^n|B^{n-1}, X^n} \sim \{P_{A_i|A^{i-1}, B^{i-1}, X^i}(\cdot|\cdot, \cdot, \cdot) : i \in \mathbb{N}^n\}$ ,  $\vec{P}_{Y^n|B^n} \sim \{P_{Y_i|Y^{i-1}, B^i}(\cdot|\cdot, \cdot) : i \in \mathbb{N}^n\}$ , is the code (i.e.,  $\{\text{encoder, decoder}\}$ ),  $\vec{P}_{B^n|A^n} \sim \{P_{B_i|B^{i-1}, A^i}(\cdot|\cdot, \cdot) : i \in \mathbb{N}^n\}$  is the channel, with excess distortion probability

$$\mathbb{P}\{d_{0,n}(X^n, Y^n) > (n+1)d\} \leq \epsilon, \quad \epsilon \in (0, 1), \quad d > 0$$

and transmission cost

$$\frac{1}{n+1} \mathbb{E}\{c_{0,n}(A^n, B^{n-1})\} \leq P, \quad P > 0$$

where  $\mathbb{P}\{\cdot\}$  and  $\mathbb{E}\{\cdot\}$  are taken with respect to the joint distribution  $P_{X^n, A^n, B^n, Y^n}(dx^n, da^n, db^n, dy^n)$  induced by  $\{\text{source, encoder, channel, decoder}\}$ .

An uncoded nonanticipative code, denoted by  $(n, d, \epsilon)$ , is a subset of an  $(n, d, \epsilon, P)$  nonanticipative code in which an encoder and decoder are identity maps,  $P_{A_i|A^{i-1}, B^{i-1}, X^i}(da_i|a^{i-1}, b^{i-1}, x^i) = \delta_{X_i}(da_i)$ ,  $P_{Y_i|Y^{i-1}, B^i}(dy_i|y^{i-1}, b^i) = \delta_{B_i}(dy_i)$ , that is,  $A_i = X_i$ ,  $Y_i = B_i$ ,  $i = 0, 1, \dots, n$ .

The well-known *Example-IID-BSS* of JSCC design utilizes an uncoded nonanticipative code, while *Example-IID-GS* utilizes an  $\{\text{encoder, decoder}\}$  pair, which scale their input (see [14]).

Next, we define the minimum excess distortion as follows.

**Definition VII.2.** (Minimum excess distortion)

The minimum excess distortion achievable by a nonanticipative code  $(n, d, \epsilon, P)$  is defined by

$$D^o(n, \epsilon, P) \triangleq \inf \left\{ d : \exists (n, d, \epsilon, P) \text{ nonanticipative code} \right\}. \quad (\text{VII.2})$$

For uncoded nonanticipative code (VII.2) is replaced by

$$\bar{D}^o(n, \epsilon) \triangleq \inf \left\{ d : \exists (n, d, \epsilon) \text{ nonanticipative code} \right\}. \quad (\text{VII.3})$$

Note that in our definition of nonanticipative code  $(n, d, \epsilon, P)$  we have assumed indirectly that the channel capacity is defined by

$$C(P) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n+1} C_{0,n}(P) \quad (\text{VII.4})$$

where

$$C_{0,n}(P) \triangleq \sup_{\{P_{A_i|A^{i-1}, B^{i-1}}(da_i|a^{i-1}, b^{i-1}) : i=0,1,\dots,n\} \in \mathcal{P}_{0,n}(P)} I(A^n \rightarrow B^n) \quad (\text{VII.5})$$

and the average power constraint is

$$\mathcal{P}_{0,n}(P) \triangleq \left\{ \{P_{A_i|A^{i-1}, B^{i-1}}(da_i|a^{i-1}, b^{i-1}) : i = 0, 1, \dots, n\} : \frac{1}{n+1} \mathbb{E}\{c_{0,n}(A^n, B^{n-1})\} \leq P \right\}.$$

Here  $I(A^n \rightarrow B^n)$  is the directed information from  $A^n$  to  $B^n$  defined by

$$I(A^n \rightarrow B^n) \triangleq \sum_{i=0}^n I(A^i; B_i|B^{i-1}). \quad (\text{VII.6})$$

Since we consider nonanticipative transmission based on Definition VII.1, we also define the notion of probabilistic realization of the optimal nonanticipative reproduction distribution corresponding to  $R^{na}(D)$  based on nonanticipative processing of information by the encoder, channel, decoder, that is, processing symbols causally, as follows (see Fig. VII.9).

**Definition VII.3.** (Probabilistic Realization)

Given a source  $\{P_{X_i|X^{i-1}}(dx_i|x^{i-1}) : i \in \mathbb{N}^n\}$ , then a channel  $\{P_{B_i|B^{i-1}, A^i}(db_i|b^{i-1}, a^i) : i \in \mathbb{N}^n\}$  is a realization of the optimal reproduction distribution  $\{Q_{Y_i|Y^{i-1}, X^i}^*(dy_i|y^{i-1}, x^i) : i \in \mathbb{N}^n\}$  corresponding to  $R_{0,n}^{na}(D)$ , if there exists a pre-channel encoder  $\{P_{A_i|A^{i-1}, B^{i-1}, X^i}(da_i|a^{i-1}, b^{i-1}, x^i) : i \in \mathbb{N}^n\}$  and a post-channel decoder  $\{P_{Y_i|Y^{i-1}, B^i}(dy_i|y^{i-1}, b^i) : i \in \mathbb{N}^n\}$  such that

$$\vec{Q}_{Y^n|X^n}^*(dy^n|x^n) = \otimes_{i=0}^n Q_{Y_i|Y^{i-1}, X^i}^*(dy_i|y^{i-1}, x^i) = \otimes_{i=0}^n Q_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i) \quad (\text{VII.7})$$

where the joint distribution from which the RHS of (VII.7) is obtained is precisely

$$P_{X^n, A^n, B^n, Y^n}(dx^n, da^n, db^n, dy^n) = \otimes_{i=0}^n P_{Y_i|Y^{i-1}, B^i}(dy_i|y^{i-1}, b^i) \otimes P_{B_i|B^{i-1}, A^i}(db_i|b^{i-1}, a^i) \\ \otimes P_{A_i|A^{i-1}, B^{i-1}, X^i}(da_i|a^{i-1}, b^{i-1}, x^i) \otimes P_{X_i|X^{i-1}}(dx_i|x^{i-1}).$$

Moreover,  $R^{na}(D)$  is realizable if in addition the realization operates with average distortion  $D$  and  $\lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n}^*) = R^{na}(D) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na}(D)$ .

The above definition of probabilistic realization is precisely the one utilized in Section V for JSCC design with respect to achieving average end-to-end distortion; it is also the one utilized to obtain the solution of the nonanticipative RDF for the multidimensional stationary Gaussian source, depicted in Fig. I.3.

Using the above definition of probabilistic realization we now prove achievability of the nonanticipative code with respect to excess distortion probability for sources with memory.

**Theorem VII.4.** (Achievability of nonanticipative code)

**Part A.** (Coded transmission)

Suppose the following conditions hold.

- (1)  $R_{0,n}^{na}(D)$  has a solution and the optimal reproduction distribution is stationary.
- (2)  $C_{0,n}(P)$  has a solution and the maximizing distribution is stationary.
- (3) The optimal stationary reproduction distribution  $\vec{Q}_{Y^n|X^n}^*(dy^n|x^n)$  given by Theorem IV.2 is realizable, and  $R^{na}(D) = \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na}(D)$  is also realizable.
- (4) For a given  $D \in [D_{min}, D_{max}]$  there exists a  $P$  such that  $R^{na}(D) = C(P)$ .

If

$$\mathbb{P}\left\{d_{0,n}(X^n, Y^n) > (n+1)d\right\} \leq \epsilon, \quad d > D \quad (\text{VII.8})$$

where  $\mathbb{P}\{\cdot\}$  is taken with respect to  $P_{Y^n, X^n}(dy^n, dx^n) = P_{X^n}(dx^n) \otimes \vec{Q}_{Y^n|X^n}^*(dy^n|x^n)$ , then there exists an  $(n, d, \epsilon, P)$  nonanticipative code.

**Part B.** (Uncoded transmission)

Suppose the following conditions hold.

- (1) Condition **Part A.** (1) holds.
- (2) The encoder and the decoder are identity maps (uncoded), and the channel  $P_{B_i|B^{i-1}, A^i}$  corresponds to  $Q_{Y_i|Y^{i-1}, X^i}$  (i.e.,  $A_i = X_i, Y_i = B_i$ ),  $i = 0, 1, \dots, n$ .
- (3) For a given  $D \in [D_{min}, D_{max}]$ , the expression  $\lim_{n \rightarrow \infty} \frac{1}{n+1} I(A^n \rightarrow B^n)$  corresponding to the optimal reproduction distribution of  $R^{na}(D)$  is finite.

If

$$\mathbb{P}\left\{d_{0,n}(X^n, Y^n) > (n+1)d\right\} \leq \epsilon, \quad d > D \quad (\text{VII.9})$$

where  $\mathbb{P}\{\cdot\}$  is taken with respect to  $P_{Y^n, X^n}(dy^n, dx^n) = P_{X^n}(dx^n) \otimes \vec{Q}_{Y^n|X^n}^*(dy^n|x^n)$ , then there exists an uncoded  $(n, d, \epsilon)$  nonanticipative code.

*Proof:* **Part A.** If conditions (1)-(3) hold then the optimal reproduction distribution is stationary, it is realizable, and this realization achieves  $R^{na}(D)$ . By (4)  $R^{na}(D) = C(P)$ . If (VII.8) is satisfied then a nonanticipative code exists. **Part B.** This is a special case of **Part A.**; by the data processing and condition **Part B.**, (3), we know that  $R^{na}(D) \leq \lim_{n \rightarrow \infty} \frac{1}{n+1} I(A^n \rightarrow B^n) < \infty$ . Hence, if (VII.9) holds, there exists an uncoded  $(n, d, \epsilon)$  nonanticipative code.  $\square$

The method described in Theorem VII.4, **Part A.**, ensures JSCC so that the channel operates at  $C(P)$ , and hence  $R^{na}(D)$  is the minimum rate of reproducing source messages at the decoder, i.e.,  $R^{na}(D) = C(P)$ . This noisy coding theorem is the one applied to the multidimensional Gaussian example, with respect to the average distortion instead of the excess distortion probability. The method described in Theorem VII.4, **Part B.**, is simpler; find the optimal reproduction distribution of  $R^{na}(D)$ , then use this distribution as the channel and ensure that (VII.9) holds, which implies achievability of the uncoded nonanticipative code. The only disadvantage is the loss of resources, because in general, the channel corresponding to the optimal reproduction distribution of  $R^{na}(D)$  will have higher capacity than the value of  $R^{na}(D)$ . With respect to the terminology in [42], this means that the source is not probabilistically matched to the channel.

Below, we apply Theorem VII.4, **Part A.** to multidimensional Gaussian process (I.11).

*Example 1: Excess Distortion Probability of Multidimensional Gaussian Process.* Consider the multidimensional stationary Gaussian source and its nonanticipative rate  $R^{na}(D)$  computed in Theorem IV.9. For a given  $D > 0$ , using coded transmission as in Theorem VII.4, **Part A.** the calculation of the excess distortion probability  $\mathbb{P}\left\{d_{0,n-1}(X^{n-1}, Y^{n-1}) > nd\right\} \leq \epsilon, \epsilon \in$



$(0, 1)$ ,  $d > D$ , can be done using Cramer's theorem [45] as follows. First note that all conditions of Theorem VII.4 are satisfied. It remains to state how to compute the excess distortion probability. By Chernoff bound we have the following.

$$\mathbb{P}\left\{\sum_{i=0}^{n-1} \rho(X_i, Y_i) > nd\right\} = \mathbb{P}\left\{\sum_{i=0}^{n-1} \rho(K_i, \tilde{K}_i) > nd\right\} \leq e^{-\lambda nd} \mathbb{E}\left\{e^{\lambda \sum_{i=0}^{n-1} \rho(K_i, \tilde{K}_i)}\right\}, \quad \lambda > 0, d > D.$$

Optimizing over all  $\lambda > 0$  then

$$\frac{1}{n} \log \mathbb{P}\left\{\sum_{i=0}^{n-1} d_{0,n-1}(X^{n-1}, Y^{n-1}) > nd\right\} \leq -\sup_{\lambda > 0} \left\{\lambda d - \frac{1}{n} \mathbb{E}\left\{e^{\lambda \sum_{i=0}^{n-1} \rho(K_i, \tilde{K}_i)}\right\}\right\}, \quad d > D.$$

Clearly, the rate function is defined by

$$I_{0,n-1}(d) = \sup_{\lambda > 0} \left\{\lambda d - \frac{1}{n} \mathbb{E}\left\{e^{\lambda \sum_{i=0}^{n-1} \rho(K_i, \tilde{K}_i)}\right\}\right\} \stackrel{(a)}{\geq} 0, \quad d > D$$

where (a) follows from the properties of the rate function (see [45, Lemma 1.2.3, p. 3]). That is, the bound for  $d > D$  is nontrivial. Indeed, for fixed  $n-1$ , it can be shown that  $I_{0,n-1}(d)$  is convex, non-decreasing function of  $d \in [D, \infty]$ . Next, we show how to compute  $\mathbb{E}\left\{e^{\lambda \sum_{i=0}^{n-1} \rho(K_i, \tilde{K}_i)}\right\}$ , when  $\rho(K_i, \tilde{K}_i) = \|K_i - \tilde{K}_i\|_{\mathbb{R}^p}$ . By (V.2), then

$$e_i = \tilde{K}_i - K_i = E_{\infty}^{tr} H_{\infty} E_{\infty} K_i + E_{\infty}^{tr} \mathcal{B}_{\infty} V_i^c - K_i = (E_{\infty}^{tr} H_{\infty} E_{\infty} - I) K_i + E_{\infty}^{tr} \mathcal{B}_{\infty} V_i^c, \quad i = 0, 1, \dots \quad (\text{VII.10})$$

By (IV.16), then

$$K_i = X_i - \hat{X}_{i|i-1} = CZ_i + NV_i - C\hat{Z}_{i|i-1} = C(Z_i - \hat{Z}_{i|i-1}) + NV_i. \quad (\text{VII.11})$$

Define the error

$$\begin{aligned} \bar{e}_{i+1} &\triangleq Z_{i+1} - \hat{Z}_{i+1|i} = AZ_i + BW_i - A\hat{Z}_{i|i-1} - A\Sigma_{\infty}(E_{\infty}^{tr} H_{\infty} E_{\infty} C)^{tr} M_{\infty}^{-1}(Y_i - C\hat{Z}_{i|i-1}) \\ &\stackrel{(b)}{=} A\bar{e}_i + BW_i - A\Sigma_{\infty}(E_{\infty}^{tr} H_{\infty} E_{\infty} C)^{tr} M_{\infty}^{-1}(E_{\infty}^{tr} H_{\infty} E_{\infty} C\bar{e}_i + E_{\infty}^{tr} H_{\infty} E_{\infty} NV_i + E_{\infty}^{tr} \mathcal{B}_{\infty} V_i^c) \\ &= (A - A\Sigma_{\infty}(E_{\infty}^{tr} H_{\infty} E_{\infty} C)^{tr} M_{\infty}^{-1} E_{\infty}^{tr} H_{\infty} E_{\infty} C)\bar{e}_i + BW_i - A\Sigma_{\infty}(E_{\infty}^{tr} H_{\infty} E_{\infty} C)^{tr} M_{\infty}^{-1} E_{\infty}^{tr} \mathcal{B}_{\infty} (\mathcal{A}_{\infty} E_{\infty} NV_i + V_i^c) \\ &\triangleq \tilde{A}\bar{e}_i + \tilde{B}_1 W_i + \tilde{B}_2 V_i + \tilde{B}_3 V_i^c, \quad \bar{e}_0 = Z_0 - \hat{Z}_{0|-1} \end{aligned} \quad (\text{VII.12})$$

where (b) follows from (IV.30) and  $\tilde{B}_2 = A\Sigma_{\infty}(E_{\infty}^{tr} H_{\infty} E_{\infty} C)^{tr} M_{\infty}^{-1} E_{\infty}^{tr} H_{\infty} E_{\infty} N$ ,  $\tilde{B}_3 = A\Sigma_{\infty}(E_{\infty}^{tr} H_{\infty} E_{\infty} C)^{tr} M_{\infty}^{-1} E_{\infty}^{tr} \mathcal{B}_{\infty} V_i^c$ ,  $\tilde{B}_1 = B$ ,  $\tilde{A} = A - A\Sigma_{\infty}(E_{\infty}^{tr} H_{\infty} E_{\infty} C)^{tr} M_{\infty}^{-1} E_{\infty}^{tr} H_{\infty} E_{\infty} C$ .

Clearly,  $\{\bar{e}_i : i = 0, 1, \dots\}$  is a Gaussian Markov process. Finally, the computation of  $\mathbb{E}\left\{e^{\lambda \sum_{i=0}^{n-1} \rho(K_i, \tilde{K}_i)}\right\}$  is as follows.

$$\begin{aligned} \mathbb{E}\left\{e^{\lambda \sum_{i=0}^{n-1} \rho(K_i, \tilde{K}_i)}\right\} &\stackrel{(c)}{=} \mathbb{E}\left\{e^{\lambda \sum_{i=0}^{n-1} \|(E_{\infty}^{tr} H_{\infty} E_{\infty} - I)K_i + E_{\infty}^{tr} \mathcal{B}_{\infty} V_i^c\|_{\mathbb{R}^p}^2}\right\} \\ &\stackrel{(d)}{=} \mathbb{E}\left\{e^{\lambda \sum_{i=0}^{n-1} \|(E_{\infty}^{tr} H_{\infty} E_{\infty} - I)(C\bar{e}_i + NV_i) + E_{\infty}^{tr} \mathcal{B}_{\infty} V_i^c\|_{\mathbb{R}^p}^2}\right\} \end{aligned} \quad (\text{VII.13})$$

where (c) follows due to (VII.10), (d) follows by (VII.11), and  $\bar{e}_{i+1}$  is given by (VII.12). The expectation in (VII.13) can be computed explicitly due to Gaussianity of  $\{\bar{e}_i : i = 0, 1, \dots\}$  and its independence of the Gaussian processes  $\{V_i : i = 0, 1, \dots\}$ ,  $\{V_i^c : i = 0, 1, \dots\}$ . A simple procedure to compute the continuous time analogue of (VII.13) is given in [46, Section III.B]. For any of the JSCC design of scalar sources, the above calculation can be easily carried out.

Below we consider the BSMS( $p$ ) and we compute the excess distortion probability, when the source is not probabilistically matched to the channel, using uncoded transmission (sub-optimal), ensuring reliable communication. This observation is also pointed out in [14] for the two well-known examples of JSCC design, the *Example-IID-BSS* and the *Example-IID-GS*. Recently, in [47] the JSCC design of the BSMS( $p$ ) is shown over a unit memory channel (with certain symmetry) with transmission cost, with and without feedback encoding. We do not present the optimal JSCC design, due to space limitations. However, the calculations presented below apply directly to the JSCC design in [47].

*Example 2: Excess Distortion Probability of BSMS( $p$ ).* Consider the BSMS( $p$ ) and its nonanticipative rate  $R^{na}(D)$  computed in Theorem IV.8. For a given  $D > 0$ , using uncoded transmission as in Theorem VII.4, **Part B**, the exact calculation of the excess distortion probability  $\mathbb{P}\left\{d_{0,n-1}(X^{n-1}, Y^{n-1}) > nd\right\} \leq \epsilon$ ,  $\epsilon \in (0, 1)$ ,  $d > D$ , is not as straightforward as it in the case of the IID Bernoulli source [24]. However, if we can show that the joint process  $\{(X_i, Y_i) : i = 0, 1, \dots\}$  induced by the optimal reproduction distribution and the BSMS( $p$ ) with alphabet  $\bar{\Sigma} \triangleq \{(x, y) : x \in \{0, 1\}, y \in \{0, 1\}\}$  is Markov, then we

can find an upper bound for the excess distortion probability, for finite but large enough  $n$  so the bound is non-trivial. Express the distortion function as follows.

$$z_i \triangleq (x_i, y_i), \quad S_n \triangleq \sum_{i=0}^{n-1} f(z_i), \quad f(z) \triangleq \rho(x, y) = x \oplus y, \quad i = 0, 1, \dots, n-1$$

Introduce the mapping  $\phi : \bar{\Sigma} \mapsto \Sigma \triangleq \{1, 2, 3, 4\}$ ,  $(x, y) \in \bar{\Sigma}$  such that under  $\phi$ ,  $(0, 0) \mapsto 1, (0, 1) \mapsto 2, (1, 0) \mapsto 3, (1, 1) \mapsto 4$ .

**Theorem VII.5.** *The joint process  $\{Z_i \triangleq (X_i, Y_i) : i = 0, 1, \dots\}$  generated by the optimal reproduction distribution  $P_{Y_i|Y_{i-1}, X_i}^*(y_i|y_{i-1}, x_i)$  and the BSMS( $p$ ),  $P_{X_i|X_{i-1}}(x_i|x_{i-1})$  is Markov, that is,  $P_{Z_i|Z_{i-1}}^*(z_i|z_{i-1}) = P_{Z_i|Z_{i-1}}^*(z_i|z_{i-1}), i = 0, 1, \dots$ , and its transition probability matrix denoted by  $\mathbf{\Pi} = \{\pi(i, j) \equiv P_{Z_i|Z_{i-1}}^*(i|j) : (i, j) \in \Sigma \times \Sigma\}$  is given by*

$$\mathbf{\Pi} = \begin{bmatrix} \alpha(1-p) & \beta(1-p) & \alpha p & \beta p \\ (1-\alpha)(1-p) & (1-\beta)(1-p) & (1-\alpha)p & p(1-\beta) \\ (1-\beta)p & (1-\alpha)p & (1-\beta)(1-p) & (1-\alpha)(1-p) \\ \beta p & \alpha p & \beta(1-p) & \alpha(1-p) \end{bmatrix} \quad (\text{VII.14})$$

where each column  $\{\pi(\cdot, j) : j \in \{1, 2, 3, 4\}\}$  is a probability vector, where  $\alpha$  and  $\beta$  are defined in (IV.11).

*Proof:* Recall the solution of the BSMS( $p$ ) in Theorem IV.8. Then for  $i = 1, 2, \dots$ , we have

$$\begin{aligned} P_{X_i, Y_i|X^{i-1}, Y^{i-1}}^*(x_i, y_i|x^{i-1}, y^{i-1}) &= Q_{Y_i|Y^{i-1}, X^i}^*(y_i|y^{i-1}, x^i) P_{X_i|X^{i-1}, Y^{i-1}}(x_i|x^{i-1}, y^{i-1}) \\ &= Q_{Y_i|Y_{i-1}, X_i}^*(y_i|y_{i-1}, x_i) P_{X_i|X_{i-1}}(x_i|x_{i-1}). \\ &= P_{X_i, Y_i|X_{i-1}, Y_{i-1}}^*(x_i, y_i|x_{i-1}, y_{i-1}). \end{aligned}$$

This shows that the joint process is Markov. By simple algebra, we obtain (VII.14).  $\square$

Since the process  $\{Z_i : i = 0, 1, \dots\}$  is Markov we can apply Hoeffding's inequality [48], which bounds the probability of a function of the Markov chain to obtain an upper bound for the excess distortion probability as follows.

$$\mathbb{P}^*\left(\frac{S_n - \mathbb{E}[S_n]}{n} \geq \gamma\right) \leq \exp\left(-\frac{\lambda^2(n\gamma - 2\|f\|m/\lambda)^2}{2n\|f\|^2 m^2}\right), \quad n > 2\|f\|m/(\lambda\gamma) \quad (\text{VII.15})$$

where  $\|f\| = 1$ ,  $m = 1$  and  $\lambda = \min\{p, 1-p\} \min\{\alpha, \beta, 1-\alpha, 1-\beta\}$ . Moreover, since  $\mathbf{\Pi}$  defined by (VII.14) is a Reversible Finite State Markov Chain (R-FSMC), a tighter bound is obtained by applying [49], and it is given by

$$\mathbb{P}^*\left(\frac{S_n - \mathbb{E}[S_n]}{n} \geq \gamma\right) \leq \exp\left(-2\frac{1-\lambda_0}{1+\lambda_0} n\gamma^2\right) \quad (\text{VII.16})$$

where  $\lambda_0 = \max(0, \lambda_2)$ , and  $\lambda_2$  is the second largest eigenvalue of the transition probability matrix  $\mathbf{\Pi}$ . The fact that  $\mathbf{\Pi}$  is R-FSMC is easily shown by verifying that  $\{\pi(i, j) : (i, j) \in \Sigma \times \Sigma\}$  and its steady state probability  $\{\pi(i) : i \in \Sigma\}$  satisfy  $\pi(i)\pi(j, i) = \pi(j)\pi(i, j)$ ,  $\forall i, j$ , or by applying Kolmogorov's necessary and sufficient criterion of reversibility given by [50]

$$\pi(j_1, j_2)\pi(j_2, j_3) \dots \pi(j_n, j_1) = \pi(j_1, j_n) \dots \pi(j_3, j_2)\pi(j_2, j_1).$$

This bound is illustrated in Fig. VII.10,  $p = 0.3$ ,  $D = 0.1$ ,  $\gamma = 0.1$ , and illustrates how the upper bound (VII.16) of the excess distortion probability changes as a function of the number of transmissions. Next, we use results from large deviations theory to compute the error exponent of the excess distortion probability, as  $n \rightarrow \infty$ . Let  $\mathbb{P}_\sigma^\pi$  denote the joint probability associated with  $\mathbf{\Pi}$ , having an initial state  $Z_0 = \sigma \in \Sigma$  given by

$$\mathbb{P}_\sigma^\pi(Z_1 = z_1, \dots, Z_{n-1} = z_{n-1}) = \pi(z_1, \sigma) \otimes_{i=2}^{n-1} \pi(z_i, z_{i-1})$$

and denote the expectation with respect to  $\mathbb{P}_\sigma^\pi$  by  $\mathbb{E}_\sigma^\pi(\cdot)$ . It can be verified that  $\mathbf{\Pi}$  is irreducible, therefore we can describe the error exponent of the excess distortion probability by the Perron-Frobenius eigenvalue of a certain non-negative matrix [51].

Let  $\lambda \in \mathbb{R}$  and define the non-negative matrix  $\mathbf{\Pi}_\lambda$  with eigenvalues

$$\pi_\lambda(j, i) = \pi(j, i) \exp\{\lambda f(j)\}, \quad i, j \in \Sigma.$$

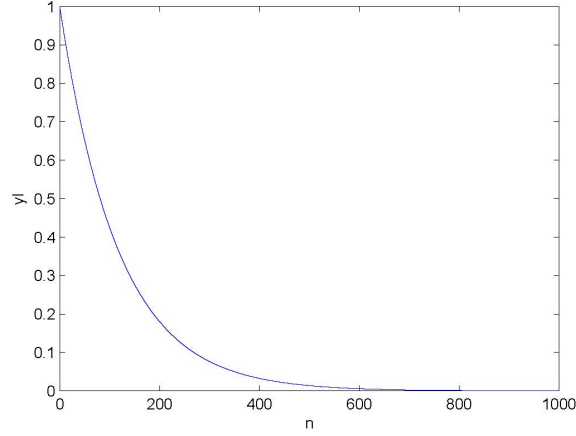


Fig. VII.10: Excess distortion probability bound (VII.16) for  $p = 0.3$ ,  $D = 0.1$  and  $\gamma = 0.1$  ( $d = D + \gamma$ ).

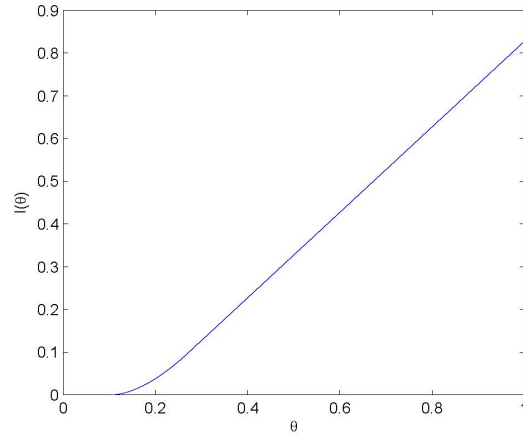


Fig. VII.11: Rate function (VII.17) for  $p = 0.3$  and  $D = 0.1$ .

Let  $\rho(\Pi_\lambda)$  denote the Perron-Frobenius eigenvalue of  $\Pi_\lambda$ . Then by [51, Theorem 3.1.2], we have the following. For any set  $\Gamma \in \mathbb{R}$  and initial state  $\sigma \in \Sigma$ ,

$$-\inf_{\theta \in \Gamma^\circ} I(\theta) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\sigma^\pi \left( \frac{S_n}{n} \in \Gamma \right) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\sigma^\pi \left( \frac{S_n}{n} \in \Gamma \right) \leq -\inf_{\theta \in \bar{\Gamma}} I(\theta)$$

where  $\Gamma^\circ$  is the interior of  $\Gamma$  and  $\bar{\Gamma}$  is the closure of  $\Gamma$ , and  $I : \mathbb{R} \mapsto [0, \infty]$  is a convex rate function defined by

$$I(\theta) \triangleq \sup_{\lambda \in \mathbb{R}} \left\{ \lambda \theta - \log \rho(\Pi_\lambda) \right\}. \quad (\text{VII.17})$$

As  $n$  becomes very large (i.e.,  $n \rightarrow \infty$ ) then

$$\mathbb{P}_\sigma^\pi \left( \frac{S_n}{n} \geq d \right) \sim \exp \left( -n \inf_{\theta \in [d, \infty]} I(\theta) \right), \quad d \geq D \quad (\text{VII.18})$$

and the exponential decay of this probability is obtained by minimizing the rate function over  $\theta \in [d, \infty]$ . Since for the evaluation of (VII.18),  $I(\theta)$  is convex, non-decreasing function of  $\theta \in [d, \infty]$ , then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\sigma^\pi \left( \frac{S_n}{n} \geq d \right) = -I(d).$$

The graph of  $I(\theta)$ , for  $\theta \geq D$  is shown in Fig. VII.11, and indicates the rate of exponential decay of the excess distortion probability as a function of  $\theta \geq D$ .

### B. Noiseless Coding Theorem

It is straight forward to verify by invoking Lemma III.2 and Theorem III.3, that the coding theorem derived in [8, Chapter 5] for two dimensional sources  $X^{n,s} \triangleq \{X_{i,j} : i = 0, \dots, n, j = 0, \dots, s\} \in \times_{i=0}^n \times_{j=0}^s \mathcal{X}_{i,j}$ , where  $i$  represents time index and  $j$  represents spatial index, gives an operational meaning to  $R_{0,n}^{na}(D)$  (even for finite  $n$ , when the process is IID in the spatial component). The connection of the coding theorem for the so-called sequential RDF derived in [8, Chapter 5] and nonanticipative RDF is described in [52, Appendix G]. The aforementioned coding theorem for two-dimensional sources is recently utilized in video coding applications [53], where the authors derived a coding theorem giving an operational meaning to  $R_{0,n}^{na}(D)$ .

### VIII. CONCLUSION

A variant of the classical RDF of the OPTA by noncausal codes, called information nonanticipative RDF, which imposes a nonanticipative or causality constraint on the optimal reproduction conditional distribution is investigated in the context of its applications in JSCC design using nonanticipative transmission schemes with respect to average and excess distortion probability, and in evaluating the RL of the OPTA by zero-delay and causal codes with respect to noncausal codes. These applications are employed to two working examples, the BSMS( $p$ ) with Hamming distance distortion, and the multidimensional partially observed Gaussian-Markov source. It is our belief that the results derived in this paper provide a crucial step towards the complete investigation of two fundamental problems in information theory, the evaluation of nonanticipative RDF in systems where sources with memory are considered, and in nonanticipative JSCC system design.

### ACKNOWLEDGEMENT

The authors wish to acknowledge the contribution of the associate editor and the reviewers in improving the presentation and results of the paper.

### APPENDIX A PROOF OF LEMMA III.2

The equivalence of **MC1**, **MC2** and **MC3** is straightforward hence it is omitted. To this end, we show equivalence of **MC4** to any of **MC1**, **MC2** and **MC3**. We proceed with the derivation, by often assuming existence of densities, which are denoted by lower case letters  $\bar{p}(\cdot|\cdot)$ , to avoid lengthy measure theoretic arguments.

**MC4**  $\implies$  **MC3**: Since for  $i = 0, \dots, n-1$ , by **MC4** we have

$$P_{X_{i+1}^n|X^i,Y^i}(dx_{i+1}^n|x^i,y^i) = P_{X_{i+1}^n|X^i}(dx_{i+1}^n|x^i)$$

then by integrating over  $\mathcal{X}_{i+2,n}$  both sides of the previous identity we obtain **MC3**.

**MC4**  $\Leftarrow$  **MC3**: Since **MC3**  $\iff$  **MC2**, we show that if  $X_{i+1}^n \leftrightarrow (X^i, Y^{i-1}) \leftrightarrow Y_i$  forms a MC for  $i = 0, 1, \dots, n-1$ , then  $X_{i+1}^n \leftrightarrow X^i \leftrightarrow Y^i$  forms a MC for  $i = 0, 1, \dots, n-1$ . We show this by induction. First, we show that  $(X_{i+1}, X_{i+2}) \leftrightarrow X^i \leftrightarrow Y^i$  forms a MC, or equivalently,  $\bar{p}(x_{i+1}, x_{i+2}|x^i, y^i) = \bar{p}(x_{i+1}, x_{i+2}|x^i)$ . Since

$$\begin{aligned} \bar{p}(x_{i+1}, x_{i+2}|x^i, y^i) &= \frac{\bar{p}(x^i, x_{i+1}, x_{i+2}, y^i)}{\bar{p}(x^i, y^i)} = \frac{\bar{p}(y_i|y^{i-1}, x^{i+2})\bar{p}(y^{i-1}, x^{i+2})}{\bar{p}(x^i, y^i)} \\ &= \frac{\underbrace{\bar{p}(y_i|y^{i-1}, x^i)}_{(a)}\bar{p}(x_{i+2}|x^{i+1}, y^{i-1})\bar{p}(x^{i+1}, y^{i-1})}{\bar{p}(x^i, y^i)} \\ &= \frac{\bar{p}(y_i|y^{i-1}, x^i)\underbrace{\bar{p}(x_{i+2}|x^{i+1})}_{(b)}\bar{p}(x_{i+1}|x^i, y^{i-1})\bar{p}(x^i, y^{i-1})}{\bar{p}(y_i|y^{i-1}, x^i)\bar{p}(x^i, y^{i-1})} \\ &= \bar{p}(x_{i+2}|x^{i+1})\underbrace{\bar{p}(x_{i+1}|x^i)}_{(c)} = \bar{p}(x_{i+2}, x_{i+1}|x^i) \end{aligned}$$

where (a) is implied from **MC2**, while (b), (c) follows from **MC3**  $\iff$  **MC2**. Hence, **MC4** holds for  $n = i+2$ . Suppose  $X_{i+1}^k \leftrightarrow X^i \leftrightarrow Y^i$  forms a MC, for some  $i+2 \leq k < n-1$ . We show that it holds for  $k \rightarrow k+1$ .

$$\begin{aligned} \bar{p}(x_{i+1}^{k+1}|x^i, y^i) &= \frac{\bar{p}(x_{i+1}^{k+1}, x^i, y^i)}{\bar{p}(x^i, y^i)} = \frac{\bar{p}(x_{k+1}|x_{i+1}^k, x^i, y^i)\bar{p}(x_{i+1}^k, x^i, y^i)}{\bar{p}(x^i, y^i)} \\ &= \frac{\underbrace{\bar{p}(x_{k+1}|x^k)}_{(d)}\bar{p}(x_{i+1}^k|x^i, y^i)\bar{p}(x^i, y^i)}{\bar{p}(x^i, y^i)} = \bar{p}(x_{k+1}|x^k)\underbrace{\bar{p}(x_{i+1}^k|x^i)}_{(e)} = \bar{p}(x_{i+1}^{k+1}|x^i) \end{aligned}$$

where (d), (e) follow from **MC3**  $\iff$  **MC2**. This completes the derivation.  $\square$

## APPENDIX B PROOF OF THEOREM IV.6

Let  $s$  be the the Lagrange multiplier which is part of the optimal solution which solves the information nonanticipative RDF. Then, by Theorem IV.2 the optimal reproduction distribution is expressed as

$$Q_{Y_i|Y^{i-1}, X^i}^*(F_i|y^{i-1}, x^i) = \int_{F_i} e^{s\rho(T^i x^n, T^i y^n)} \lambda_i(x^i, y^{i-1}) P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1}), \quad \forall F_i \in \mathcal{B}(\mathcal{Y}_i), \quad \forall i \in \mathbb{N}^n. \quad (\text{B.1})$$

By integrating (B.1) with respect to  $P_{0,i}^*(dx^i|y^{i-1})$  we obtain the expression

$$\begin{aligned} \int_{\mathcal{X}_{0,i}} Q_{Y_i|Y^{i-1}, X^i}^*(F_i|y^{i-1}, x^i) \otimes P_{X^i|Y^{i-1}}^*(dx^i|y^{i-1}) &= P_{X^i, Y_i|Y^{i-1}}^*(\mathcal{X}_{0,i} \times F_i|y^{i-1}) \\ &= \int_{F_i \times \mathcal{X}_{0,i}} e^{s\rho(T^i x^n, T^i y^n)} \lambda_i(x^i, y^{i-1}) P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1}) \otimes P_{X^i|Y^{i-1}}^*(dx^i|y^{i-1}), \quad \forall F_i \in \mathcal{B}(\mathcal{Y}_i), \quad i \in \mathbb{N}^n. \end{aligned}$$

Moreover,  $\forall F_i \in \mathcal{B}(\mathcal{Y}_i), \quad i \in \mathbb{N}^n$ ,

$$\begin{aligned} P_{Y_i|Y^{i-1}}^*(F_i|y^{i-1}) &= \int_{F_i} P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1}) = P_{X^i, F_i|Y^{i-1}}^*(\mathcal{X}_{0,i} \times F_i|y^{i-1}) \\ &= \int_{F_i \times \mathcal{X}_{0,i}} e^{s\rho(T^i x^n, T^i y^n)} \lambda_i(x^i, y^{i-1}) P_{0,i}^*(dx^i|y^{i-1}) \otimes P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1}). \end{aligned} \quad (\text{B.2})$$

Utilizing (B.2) we finally obtain

$$\int_{\mathcal{X}_{0,i}} e^{s\rho(T^i x^n, T^i y^n)} \lambda_i(x^i, y^{i-1}) P_{X^i|Y^{i-1}}^*(dx^i|y^{i-1}) = 1, \quad P_{Y_i|Y^{i-1}}^* - a.s., \quad \forall i \in \mathbb{N}.$$

This completes the derivation.  $\square$

## APPENDIX C PROOF OF THEOREM IV.7

Let  $s \leq 0$ ,  $\lambda \in \Psi_s$  and  $\vec{Q}_{Y^n|X^n}(\cdot|x^n) \in \vec{\mathcal{Q}}_{0,n}(D)$  be given. Then, using the fact that

$$\frac{1}{n+1} \sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i}} \rho(T^i x^n, T^i y^n) (P_{X^i} \otimes \vec{Q}_{Y^i|X^i})(dx^i, dy^i) \leq D$$

gives

$$\begin{aligned} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \vec{Q}_{Y^n|X^n}) &- sD(n+1) - \sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i-1}} \log \left( \lambda_i(x^i, y^{i-1}) \right) (P_{X^i} \otimes \vec{Q}_{Y^{i-1}|X^{i-1}})(dx^i, dy^{i-1}) \\ &\geq \sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i}} \log \left( \frac{Q_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i)}{P_{Y_i|Y^{i-1}}(dy_i|y^{i-1})} \right) (P_{X^i} \otimes \vec{Q}_{Y^i|X^i})(dx^i, dy^i) \\ &\quad - s \sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i}} \rho(T^i x^n, T^i y^n) (P_{X^i} \otimes \vec{Q}_{Y^i|X^i})(dx^i, dy^i) \\ &\quad - \sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i}} \log \left( \lambda_i(x^i, y^{i-1}) \right) (P_{X^i} \otimes \vec{Q}_{Y^i|X^i})(dx^i, dy^i) \\ &= \sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i}} \log \left( \frac{Q_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i) e^{-s\rho(T^i x^n, T^i y^n)}}{P_{Y_i|Y^{i-1}}(dy_i|y^{i-1}) \lambda_i(x^i, y^{i-1})} \right) (P_{X^i} \otimes \vec{Q}_{Y^i|X^i})(dx^i, dy^i) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=0}^n \int_{\mathcal{X}_{0,i-1} \times \mathcal{Y}_{0,i-1}} \left\{ \int_{\mathcal{X}_i \times \mathcal{Y}_i} \log \left( \frac{Q_{Y_i|Y^{i-1},X^i}(dy_i|y^{i-1},x^i) e^{-s\rho(T^i x^n, T^i y^n)}}{P_{Y_i|Y^{i-1}}(dy_i|y^{i-1}) \lambda_i(x^i, y^{i-1})} \right) \right. \\
&\quad \left. Q_{Y_i|Y^{i-1},X^i}(dy_i|y^{i-1},x^i) \otimes P_{X_i|X^{i-1}}(dx_i|x^{i-1}) \right\} \otimes P_{X^{i-1},Y^{i-1}}(dx^{i-1}, dy^{i-1}) \\
&\stackrel{(a)}{\geq} \sum_{i=0}^n \int_{\mathcal{X}_{0,i-1} \times \mathcal{Y}_{0,i-1}} \left\{ \int_{\mathcal{X}_i \times \mathcal{Y}_i} \left( 1 - \frac{e^{s\rho(T^i x^n, T^i y^n)} P_{Y_i|Y^{i-1}}(dy_i|y^{i-1}) \lambda_i(x^i, y^{i-1})}{Q_{Y_i|Y^{i-1},X^i}(dy_i|y^{i-1},x^i)} \right) \right. \\
&\quad \left. Q_{Y_i|Y^{i-1},X^i}(dy_i|y^{i-1},x^i) \otimes P_{X_i|X^{i-1}}(dx_i|x^{i-1}) \right\} \otimes P_{X^{i-1},Y^{i-1}}(dx^{i-1}, dy^{i-1}) \\
&= \sum_{i=0}^n \left\{ 1 - \int_{\mathcal{X}_{0,i-1} \times \mathcal{Y}_{0,i-1}} \int_{\mathcal{X}_i \times \mathcal{Y}_i} e^{s\rho(T^i x^n, T^i y^n)} \lambda_i(x^i, y^{i-1}) P_{Y_i|Y^{i-1}}(dy_i|y^{i-1}) \right. \\
&\quad \left. \otimes P_{X_i|X^{i-1}}(dx_i|x^{i-1}) \otimes P_{X^{i-1},Y^{i-1}}(dx^{i-1}, dy^{i-1}) \right\} \\
&= \sum_{i=0}^n \left\{ 1 - \int_{\mathcal{Y}_i} P_{Y_i|Y^{i-1}}(dy_i|y^{i-1}) \int_{\mathcal{X}_{0,i-1}} P_{Y^{i-1}}(dy^{i-1}) \left( \int_{\mathcal{X}_{0,i}} e^{s\rho(T^i x^n, T^i y^n)} \lambda_i(x^i, y^{i-1}) \otimes P_{X^i|Y^{i-1}}(dx^i|y^{i-1}) \right) \right\} \\
&\stackrel{(b)}{\geq} \sum_{i=0}^n \left( 1 - \int_{\mathcal{Y}_{0,i}} P_{Y^i}(dy^i) \right) = 0
\end{aligned}$$

where (a) follows from the inequality  $\log x \geq 1 - \frac{1}{x}$ ,  $x > 0$ , and (b) follows from (IV.9). Hence,

$$R_{0,n}^{na}(D) \stackrel{(c)}{\geq} \sup_{s \leq 0} \sup_{\lambda \in \Psi_s} \left\{ sD(n+1) + \sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i-1}} \log \left( \lambda_i(x^i, y^{i-1}) \right) P_{X^{i-1},Y^{i-1}}(dx^{i-1}, dy^{i-1}) \otimes P_{X_i|X^{i-1}}(dx_i|x^{i-1}) \right\}.$$

However, equality in (c) holds if

$$\lambda_i(x^i, y^{i-1}) \triangleq \left( \int_{\mathcal{Y}_i} e^{s\rho(T^i x^n, T^i y^n)} P_{Y_i|Y^{i-1}}^*(dy_i|y^{i-1}) \right)^{-1}, \quad i = 0, 1, \dots, n.$$

This completes the derivation.  $\square$

#### APPENDIX D PROOF OF THEOREM IV.9

The derivation is based on showing that  $R^{na}(D)$  is bounded above and below by the RHS of (IV.20). The lower bound is obtained by using Theorem IV.7, to derive a lower bound analogous to Shannon's lower bound. Define

$$H_\infty = \lim_{t \rightarrow \infty} H_t, \quad H_t \triangleq \text{diag}\{\eta_{t,1}, \dots, \eta_{t,p}\}, \quad \eta_{t,i} = 1 - \frac{\delta_{t,i}}{\lambda_{t,i}}, \quad i = 1, \dots, p, \quad t \in \mathbb{N}.$$

Consider the additive noisy channel with feedback of the form

$$\tilde{K}_t = E_t^{tr} H_t E_t (X_t - \mathbb{E}\{X_t | \sigma\{Y^{t-1}\}\}) + E_t^{tr} \mathcal{B}_t V_t^c = E_t^{tr} H_t E_t K_t + E_t^{tr} \mathcal{B}_t V_t^c, \quad t \in \mathbb{N} \quad (\text{D.1})$$

where  $\{V_t^c : t \in \mathbb{N}\}$  is an independent Gaussian zero mean process with covariance  $\text{cov}(V_t^c) = Q = \text{diag}\{q_1, \dots, q_p\}$ , and  $\{\mathcal{B}_t : t \in \mathbb{N}\}$  is to be determined.

Next, we show by letting  $\mathcal{B}_\infty = \lim_{t \rightarrow \infty} \mathcal{B}_t$ , where  $\mathcal{B}_t = \sqrt{H_t \Delta_t Q^{-1}}$ , and  $\Delta_t \triangleq \text{diag}\{\delta_{t,1}, \dots, \delta_{t,p}\}$ , that  $\Lambda_\infty = \lim_{t \rightarrow \infty} \Lambda_t = \lim_{t \rightarrow \infty} \mathbb{E}\{K_t K_t^{tr}\}$ , and also  $\lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{E}\left\{ \sum_{t=0}^n \|X_t - Y_t\|_{\mathbb{R}^p}^2 \right\} = \lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{E}\left\{ \sum_{t=0}^n \|K_t - \tilde{K}_t\|_{\mathbb{R}^p}^2 \right\} = D$ . Clearly, by (IV.16), (IV.18), (D.1)

$$\begin{aligned}
&\lim_{t \rightarrow \infty} \mathbb{E}\{(X_t - Y_t)^{tr} (X_t - Y_t)\} = \lim_{t \rightarrow \infty} \text{Trace} \mathbb{E}\{(K_t - \tilde{K}_t)(K_t - \tilde{K}_t)^{tr}\} \\
&= \lim_{t \rightarrow \infty} \text{Trace} \mathbb{E}\{(K_t - E_t^{tr} H_t E_t K_t - E_t^{tr} \mathcal{B}_t V_t^c)(K_t - E_t^{tr} H_t E_t K_t - E_t^{tr} \mathcal{B}_t V_t^c)^{tr}\} \\
&= \lim_{t \rightarrow \infty} \text{Trace} \left\{ E_t^{tr} \left( (I - H_t) \text{diag}(\lambda_{t,1}, \dots, \lambda_{t,p}) (I - H_t)^{tr} + (\mathcal{B}_t Q \mathcal{B}_t^{tr}) \right) E_t \right\} \\
&\stackrel{(a)}{=} \lim_{t \rightarrow \infty} \text{Trace} \left\{ \text{diag}(\delta_{t,1}, \dots, \delta_{t,p}) \right\} = \text{Trace} \left\{ \text{diag}(\delta_{\infty,1}, \dots, \delta_{\infty,p}) \right\} = D
\end{aligned}$$

where (a) holds by setting  $\mathcal{B}_\infty$  and  $\mathcal{B}_t$  as in (IV.27). Also, by (IV.19),

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na,K^n,\tilde{K}^n}(D) &\leq \lim_{n \rightarrow \infty} \frac{1}{n+1} \mathbb{I}_{X^n \rightarrow Y^n}(P_{K^n}, \vec{Q}_{\tilde{K}^n|K^n}) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n I(K_t; \tilde{K}_t | \tilde{K}^{t-1}) = \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \left( H(\tilde{K}_t | \tilde{K}^{t-1}) - H(\tilde{K}_t | \tilde{K}^{t-1}, K_t) \right) \\
&\stackrel{(b)}{\leq} \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \left( H(\tilde{K}_t) - H(\tilde{K}_t | \tilde{K}^{t-1}, K_t) \right) \stackrel{(c)}{\leq} \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \left( H(\tilde{K}_t) - H(\tilde{K}_t | K_t) \right) \\
&\stackrel{(d)}{\leq} \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \left( H(\tilde{K}_t) - H(E_t^{tr} \mathcal{B}_t V_t^c) \right)
\end{aligned} \tag{D.2}$$

where (b) follows from the fact that conditioning reduces entropy, (c) follows from the fact that  $\tilde{K}_t = E_t^{tr} H_t E_t K_t + E_t^{tr} \mathcal{B}_t V_t^c$  is a memoryless Gaussian channel, and (d) follows from the orthogonality of  $K_t$  and  $V_t^c$ ,  $\forall t \in \mathbb{N}$ . Next, we compute the entropy rates appearing in (D.2). The covariance of the Gaussian zero mean, noise process  $\{E_t^{tr} \mathcal{B}_t V_t^c, t \in \mathbb{N}\}$  is obtained as follows.

$$\begin{aligned}
\lim_{t \rightarrow \infty} \mathbb{E} \left\{ (E_t^{tr} \mathcal{B}_t V_t^c) (E_t^{tr} \mathcal{B}_t V_t^c)^{tr} \right\} &= \lim_{t \rightarrow \infty} \mathbb{E} \left\{ E_t^{tr} \mathcal{B}_t V_t^c V_t^{c,tr} \mathcal{B}_t^{tr} E_t \right\} \\
&= \lim_{t \rightarrow \infty} \left\{ E_t^{tr} \sqrt{H_t \Delta_t Q^{-1}} Q \sqrt{H_t \Delta_t Q^{-1}} E_t \right\} \\
&= \lim_{t \rightarrow \infty} E_t^{tr} H_t \Delta_t E_t = \lim_{t \rightarrow \infty} E_t^{tr} \text{diag} \{ \eta_{t,1} \delta_{t,1}, \dots, \eta_{t,p} \delta_{t,p} \} E_t \\
&= E_\infty^{tr} \text{diag} \{ \eta_{\infty,1} \delta_{\infty,1}, \dots, \eta_{\infty,p} \delta_{\infty,p} \} E_\infty.
\end{aligned} \tag{D.3}$$

The covariance of the process  $\{\tilde{K}_t : t \in \mathbb{N}\}$  is obtained as follows.

$$\begin{aligned}
\lim_{t \rightarrow \infty} \mathbb{E} \left\{ \tilde{K}_t \tilde{K}_t^{tr} \right\} &= \lim_{t \rightarrow \infty} \mathbb{E} \left\{ E_t^{tr} H_t E_t K_t K_t^{tr} E_t^{tr} H_t E_t + E_t^{tr} \mathcal{B}_t V_t^c V_t^{c,tr} \mathcal{B}_t^{tr} E_t \right\} \\
&= \lim_{t \rightarrow \infty} \left\{ E_t^{tr} H_t E_t \Lambda_t E_t^{tr} H_t E_t + E_t^{tr} \sqrt{H_t \Delta_t Q^{-1}} Q \sqrt{H_t \Delta_t Q^{-1}} E_t \right\} \\
&= \lim_{t \rightarrow \infty} \left\{ E_t^{tr} \left( \text{diag} \{ \eta_{t,1}^2 \lambda_{t,1}, \dots, \eta_{t,p}^2 \lambda_{t,p} \} + \text{diag} \{ \eta_{t,1} \delta_{t,1}, \dots, \eta_{t,p} \delta_{t,p} \} \right) E_t \right\} \\
&= \lim_{t \rightarrow \infty} E_t^{tr} \text{diag} \{ \lambda_{t,1} - \delta_{t,1}, \dots, \lambda_{t,p} - \delta_{t,p} \} E_t, \quad \lambda_{t,i} - \delta_{t,i} \geq 0, \forall t \\
&= E_\infty^{tr} \text{diag} \{ \lambda_{\infty,1} - \delta_{\infty,1}, \dots, \lambda_{\infty,p} - \delta_{\infty,p} \} E_\infty.
\end{aligned} \tag{D.4}$$

Using (D.4) we obtain the first term of (D.2) as follows

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n H(\tilde{K}_t) &= \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \log \left\{ (2\pi e) \times_{i=1}^p (\lambda_{t,i} - \delta_{t,i})^+ \right\} \\
&= \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p \log \left\{ (2\pi e) (\lambda_{t,i} - \delta_{t,i})^+ \right\} \\
&= \frac{1}{2} \sum_{i=1}^p \log \left\{ (2\pi e) (\lambda_{\infty,i} - \delta_{\infty,i})^+ \right\}.
\end{aligned} \tag{D.5}$$

Also, by (D.3), we obtain the second term in (D.2) as follows.

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n H(E_t^{tr} \mathcal{B}_t V_t^c) &= \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \log \left\{ (2\pi e) \text{diag} \{ \eta_{t,1} \delta_{t,1}, \dots, \eta_{t,p} \delta_{t,p} \} \right\} \\
&= \frac{1}{2} \sum_{t=0}^n \log \left\{ (2\pi e) \times_{i=1}^p (\eta_{t,i} \delta_{t,i}) \right\} = \frac{1}{2} \sum_{i=1}^p \log \left\{ (2\pi e) (\eta_{\infty,i} \delta_{\infty,i}) \right\}.
\end{aligned} \tag{D.6}$$

By using (D.5) and (D.6) in (D.2) we have the following upper bound

$$\begin{aligned}
\lim_{n \rightarrow \infty} R_{0,n}^{na,K^n,\tilde{K}^n}(D) &\leq \frac{1}{2} \sum_{i=1}^p \log \left\{ (2\pi e) (\lambda_{\infty,i} - \delta_{\infty,i})^+ \right\} - \frac{1}{2} \sum_{i=1}^p \log \left\{ (2\pi e) (\eta_{\infty,i} \delta_{\infty,i}) \right\} \\
&= \frac{1}{2} \sum_{i=1}^p \log \left\{ \frac{(\lambda_{\infty,i} - \delta_{\infty,i})^+}{\eta_{\infty,i} \delta_{\infty,i}} \right\} = \frac{1}{2} \sum_{i=1}^p \log \frac{\lambda_{\infty,i}}{\delta_{\infty,i}}
\end{aligned}$$

where  $\delta_{\infty,i} = \min\{\xi_{\infty}, \lambda_{\infty,i}\}$  and  $\sum_{i=1}^p \delta_{\infty,i} = D$ .

*Achievable Lower Bound (Analogous to Shannon's Lower Bound).* Next, we apply Theorem IV.7 to obtain a lower bound for the nonanticipative RDF  $R^{na}(D) = \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na}(D) = \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na,K^n,\tilde{K}^n}(D)$ .

Applying Theorem IV.7 to  $R_{0,n}^{na,K^n,\tilde{K}^n}(D)$ , the set  $\Psi_s$  is defined by

$$\Psi_s \triangleq \left\{ \{ \lambda_t(k_t, \tilde{k}^{t-1}) : t = 0, 1, \dots, n \} : \lambda_t(k_t, \tilde{k}^{t-1}) \geq 0, \right. \\ \left. \int e^{s\|K_t - \tilde{K}_t\|_{\mathbb{R}^p}^2} \lambda(k_t, \tilde{k}^{t-1}) \bar{p}(k_t | \tilde{k}^{t-1}) dk_t \leq 1, t = 0, 1, \dots, n \right\} \quad (\text{D.7})$$

where  $\bar{p}(k_t | \tilde{k}^{t-1})$  denotes the conditional density of  $k_t$  given  $\tilde{k}^{t-1}$ . Choose  $s \leq 0$  and take  $\lambda_t(k_t, \tilde{k}^{t-1}) \in \Psi_s$  such that

$$\lambda_t(k_t, \tilde{k}^{t-1}) = \frac{\alpha_t}{\bar{p}(k_t | \tilde{k}^{t-1})} \quad (\text{D.8})$$

for some  $\alpha_t$  not depending on  $k_t$ . Substituting (D.8) into the reduced integral inequality in (D.7) we obtain

$$\alpha_t \int e^{s\|K_t - \tilde{K}_t\|_{\mathbb{R}^p}^2} dk_t \leq 1, t = 0, 1, \dots, n.$$

By changing the variable of integration we also obtain

$$\alpha_t \int e^{s\|z_t\|_{\mathbb{R}^p}^2} dz_t = \alpha_t \sqrt{\left(-\frac{\pi}{s}\right)^p} = \alpha_t \left(-\frac{\pi}{s}\right)^{\frac{p}{2}} \leq 1, t = 0, 1, \dots, n. \quad (\text{D.9})$$

By setting  $\alpha_t = \left(-\frac{s}{\pi}\right)^{\frac{p}{2}}$ ,  $t = 0, 1, \dots, n$ , the inequality of (D.9) holds with equality. Then, by Theorem IV.7, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na,K^n,\tilde{K}^n}(D) &\geq sD + \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \int_{K_t \times \tilde{K}_{0,t-1}} \bar{p}(k_t, \tilde{k}^{t-1}) \log \left( \lambda_t(k_t, \tilde{k}^{t-1}) \right) dk_t d\tilde{k}^{t-1} \\ &\stackrel{(e)}{=} sD + \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \int_{K_t \times \tilde{K}_{0,t-1}} \bar{p}(k_t, \tilde{k}^{t-1}) \log \left( \frac{\left(-\frac{s}{\pi}\right)^{\frac{p}{2}}}{\bar{p}(k_t | \tilde{k}^{t-1})} \right) dk_t d\tilde{k}^{t-1} \\ &= sD + \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \log \left( -\frac{s}{\pi} \right)^{\frac{p}{2}} + \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n H(K_t | \tilde{K}^{t-1}) \\ &\stackrel{(f)}{=} sD + \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \log \left( -\frac{s}{\pi} \right)^{\frac{p}{2}} + \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n H(K_t) \end{aligned} \quad (\text{D.10})$$

where (e) follows from (D.8), and (f) follows from the orthogonality of  $K_t$  and  $\tilde{K}^{t-1}$ . Next, we need to find the Lagrangian multiplier “ $s$ ” so that the lower bound (D.10) equals  $\frac{1}{2} \sum_{i=1}^p \log \frac{\lambda_{\infty,i}}{\delta_{\infty,i}}$ . To this end, we need to ensure existence of some  $s < 0$  such that the following identity holds.

$$\begin{aligned} sD + \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \log \left( -\frac{s}{\pi} \right)^{\frac{p}{2}} + \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \log 2\pi e |\Lambda_t| &= \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p \log \frac{\lambda_{t,i}}{\delta_{t,i}} \\ \implies sD + \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p \log \left( -\frac{s}{\pi} \right) + \frac{1}{2} \sum_{i=1}^p \log 2\pi e (\lambda_{\infty,i}) &= \frac{1}{2} \sum_{i=1}^p \log \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} \\ \implies \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \log e^{2(n+1)s \sum_{i=1}^p \delta_{t,i}} + \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p \log \left( -\frac{s}{\pi} \right) \\ &= \frac{1}{2} \sum_{i=1}^p \log \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} - \frac{1}{2} \sum_{i=1}^p \log 2\pi e (\lambda_{\infty,i}) \\ \implies \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p \log e^{2s\delta_{t,i}} + \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p \log \left( -\frac{s}{\pi} \right) &= \frac{1}{2} \sum_{i=1}^p \log \frac{1}{2\pi e \delta_{\infty,i}} \\ \implies \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p \log e^{2s\delta_{t,i}} \left( -\frac{s}{\pi} \right) &= \frac{1}{2} \sum_{i=1}^p \log \frac{1}{2\pi e \delta_{\infty,i}} \\ \implies \frac{1}{2} \sum_{i=1}^p \log e^{2s\delta_{\infty,i}} \left( -\frac{s}{\pi} \right) &= \frac{1}{2} \sum_{i=1}^p \log \frac{1}{2\pi e \delta_{\infty,i}} \implies e^{2s\delta_{\infty,i}} \left( -\frac{s}{\pi} \right) = \frac{1}{2\pi e \delta_{\infty,i}} \implies s = -\frac{1}{2\delta_{\infty,i}} \end{aligned}$$



where  $\delta_{\infty,i} = \{\xi_{\infty}, \lambda_{\infty,i}\}$ . Now, if  $\delta_{\infty,i} = \xi_{\infty}$  then  $s = -\frac{1}{2\delta_{\infty,i}}$  and the nonanticipative RDF is bounded below by the following expression

$$\lim_{n \rightarrow \infty} R_{0,n}^{na,K^n,\tilde{K}^n}(D) \geq \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p \log \frac{\lambda_{t,i}}{\delta_{t,i}} = \frac{1}{2} \sum_{i=1}^p \log \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} \quad (\text{D.11})$$

which is the desired lower bound with  $\sum_{i=1}^p \delta_{\infty,i} = D$ . When  $\delta_{\infty,i} = \lambda_{\infty,i}$ , no encoding is performed and there is no need to prove a lower bound on  $R_{0,n}^{na,K^n,\tilde{K}^n}(D)$ . This completes the proof of (IV.20).

Next, we determine the expression of  $\Lambda_{\infty}$ . By definition,  $\Lambda_{\infty} = \lim_{t \rightarrow \infty} \Lambda_t$ , where  $\Lambda_t = \text{cov}(X_t - \mathbb{E}\{X_t | \sigma\{Y^{t-1}\}\})$ . Since  $X_t - \mathbb{E}\{X_t | \sigma\{Y^{t-1}\}\} = CZ_t + NV_t - C\mathbb{E}\{Z_t | \sigma\{Y^{t-1}\}\}$  then  $\Lambda_t = C\Sigma_t C^{tr} + NN^{tr}$ . Let  $\hat{Z}_{t|t-1} = \mathbb{E}\{Z_t | \sigma\{Y^{t-1}\}\}$ . Clearly,  $\Sigma_t \triangleq \mathbb{E}\{(Z_t - \mathbb{E}\{Z_t | Y^{t-1}\})(Z_t - \mathbb{E}\{Z_t | Y^{t-1}\})^{tr}\}$ . Moreover,  $\Lambda_{\infty} = C \lim_{t \rightarrow \infty} \Sigma_t C^{tr} + NN^{tr}$ . Therefore, to determine  $\Sigma_{\infty} \triangleq \lim_{t \rightarrow \infty} \Sigma_t$ , we need the equation of the error  $e_t \triangleq Z_t - \hat{Z}_{t|t-1}$ , hence the equation of the least-squares filter of  $Z_t$  given all the previous outputs  $Y^{t-1}$ , namely  $\hat{Z}_{t|t-1}$ . From Fig. I.3, we deduce that  $Y_t = \tilde{K}_t + \hat{X}_{t|t-1}$ , where  $\{\hat{X}_{t|t-1} : t \in \mathbb{N}\}$  is obtained from the modified Kalman filter as follows. Recall that

$$\begin{aligned} Y_t = \tilde{K}_t + \hat{X}_{t|t-1} &= E_{\infty}^{tr} H_{\infty} E_{\infty} (X_t - \hat{X}_{t|t-1}) + E_{\infty}^{tr} \mathcal{B}_{\infty} V_t^c + \hat{X}_{t|t-1} \\ &= E_{\infty}^{tr} H_{\infty} E_{\infty} (CZ_t + NV_t - C\hat{Z}_{t|t-1}) + E_{\infty}^{tr} \mathcal{B}_{\infty} V_t^c + C\hat{Z}_{t|t-1} \\ &= E_{\infty}^{tr} H_{\infty} E_{\infty} C(Z_t - \hat{Z}_{t|t-1}) + C\hat{Z}_{t|t-1} + E_{\infty}^{tr} H_{\infty} E_{\infty} NV_t + E_{\infty}^{tr} \mathcal{B}_{\infty} V_t^c \end{aligned}$$

where  $\{V_t : t \in \mathbb{N}\}$  and  $\{V_t^c : t \in \mathbb{N}\}$  are independent Gaussian vectors. Then  $\hat{Z}_{t|t-1}$  is given by the modified Kalman filter [28] (IV.23), (IV.24). Notice that the filter is ergodic with initial condition  $\hat{Z}_{0|-1} = \mathbb{E}\{Z_0 | Y^{-1}\}$  and  $\Sigma_0$  the covariance of  $Z_0 - \hat{Z}_{0|-1}$  which is Gaussian  $N(0, \Sigma_{\infty})$ .  $\square$

## REFERENCES

- [1] P. A. Stavrou, C. K. Kourtellis, and C. D. Charalambous, "Applications of information nonanticipative rate distortion function," in *IEEE International Symposium on Information Theory (ISIT)*, June 29–4 July 2014, pp. 3062–3066.
- [2] —, "Information nonanticipative rate distortion function and its applications," in *Coordination Control of Distributed Systems*, ser. Lecture Notes in Control and Information Sciences, J. H. van Schuppen and T. Villa, Eds. Springer International Publishing, 2015, vol. 456, pp. 317–324.
- [3] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [4] T. S. Han, *Information-Spectrum Methods in Information Theory*, 2nd ed. Springer-Verlag, Berlin, Heidelberg, New York, 2003.
- [5] D. Neuhoff and R. Gilbert, "Causal source codes," *IEEE Transactions on Information Theory*, vol. 28, no. 5, pp. 701–713, Sep. 1982.
- [6] T. Ericson, "A result on delay-less information transmission," in *IEEE International Symposium on Information Theory (ISIT)*, 1979.
- [7] N. Gaarder and D. Slepian, "On optimal finite-state digital transmission systems," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 167–186, 1982.
- [8] S. C. Tatikonda, "Control over communication constraints," Ph.D. dissertation, Mass. Inst. of Tech. (M.I.T.), Cambridge, MA, 2000.
- [9] T. Linder and G. Lagosi, "A zero-delay sequential scheme for lossy coding of individual sequences," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2533–2538, 2001.
- [10] E. Akyol, K. B. Viswanatha, K. Rose, and T. A. Ramstad, "On zero-delay source-channel coding," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7473–7489, Dec. 2014.
- [11] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [12] T. Berger, "Living information theory," *IEEE Information Theory Society Newsletter*, vol. 53, no. 1, pp. 6–19, Mar 2003.
- [13] J. Gollub, T., "Theoretical limitations on the transmission of data from analog sources," *IEEE Transactions on Information Theory*, vol. 11, no. 4, pp. 558–567, Oct 1965.
- [14] V. Kostina and S. Verdú, "To code or not to code: Revisited," in *IEEE Information Theory Workshop (ITW)*, Sept. 2012, pp. 5–9.
- [15] M. S. Derpich and J. Østergaard, "Improved upper bounds to the causal quadratic rate-distortion function for gaussian stationary sources," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3131–3152, May 2012.
- [16] C. D. Charalambous, P. A. Stavrou, and N. U. Ahmed, "Nonanticipative rate distortion function and relations to filtering theory," *IEEE Transactions on Automatic Control*, vol. 59, no. 4, pp. 937–952, April 2014.
- [17] C. D. Charalambous and P. A. Stavrou, "On the relation of nonanticipative rate distortion function and filtering theory," in *12th Biannual European Control Conference (ECC)*, 17–19 June 2013, pp. 1627–1632.
- [18] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback—I: No bandwidth constraint," *IEEE Trans. on Info. Theory*, vol. 12, no. 2, pp. 172–182, Apr 1966.
- [19] A. K. Gorbunov and M. S. Pinsker, "Nonanticipatory and prognostic epsilon entropies and message generation rates," *Problems of Information Transmission*, vol. 9, no. 3, pp. 184–191, July–Sept. 1973, translation from Problemy Peredachi Informatsii, vol. 9, no. 3, pp. 12–21, July–September 1973.
- [20] M. S. Pinsker and A. K. Gorbunov, "Epsilon-entropy with delay for small mean-square reproduction error," *Problems of Information Transmission*, vol. 23, no. 2, pp. 91–95, 1987, translation from Problemy Peredachi Informatsii, vol. 23, no. 2, pp. 3–8, April–June 1987.
- [21] A. K. Gorbunov and M. S. Pinsker, "Asymptotic behavior of nonanticipative epsilon-entropy for Gaussian processes," *Problems of Information Transmission*, vol. 27, no. 4, pp. 361–365, 1991, translation from Problemy Peredachi Informatsii, vol. 27, no. 4, pp. 100–104, October–December 1991.
- [22] I. Csiszár, "On an extremum problem of information theory," *Studia Scientiarum Mathematicarum Hungarica*, vol. 9, pp. 57–71, 1974.
- [23] S. Ihara, *Information theory - for Continuous Systems*. World Scientific, 1993.
- [24] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, 2012.
- [25] —, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2545–2575, 2013.

- [26] M. Johnston, E. Modiano, and Y. Polyanskiy, "Opportunistic scheduling with limited channel state information: A rate distortion approach," in *IEEE International Symposium on Information Theory (ISIT)*, June 2014, pp. 1371–1375.
- [27] T. Berger, "Explicit bounds to  $r(d)$  for a binary symmetric markov source," *IEEE Transactions on Information Theory*, vol. 23, no. 1, pp. 52–59, 1977.
- [28] P. E. Caines, *Linear Stochastic Systems*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, 1988.
- [29] T. Tanaka, K. K. Kim, P. A. Parrilo, and S. K. Mitter, "Semidefinite programming approach to gaussian sequential rate-distortion trade-offs," *arxiv.org*, 2014.
- [30] T. Linder and R. Zamir, "Causal coding of stationary sources and individual sequences with high resolution," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 662–680, Feb. 2006.
- [31] C. D. Charalambous and P. A. Stavrou, "Directed information on abstract spaces: Properties and variational equalities," *submitted to IEEE Transactions on Information Theory*, 2013. [Online]. Available: <http://arxiv.org/abs/1302.3971>
- [32] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, Inc., New York, 1997.
- [33] M. Pinsker, *Information and Information Stability of Random Variables and Processes*. Holden-Day Inc, San Francisco, 1964, translated by Amiel Feinstein.
- [34] V. Solo, "On causality and mutual information," in *47th IEEE Conference on Decision and Control (CDC '08)*, Dec. 2008, pp. 4939–4944.
- [35] F. H. Lin, K. Hara, V. Solo, M. Vangel, J. W. Belliveau, S. T. Stufflebeam, and H. M. S., "Dynamic Granger-Geweke causality modeling with application to interictal spike propagation," *Human Brain Mapping*, vol. 30, no. 6, pp. 1877–1886, June 2009.
- [36] J. Pearl, *Causality: models, reasoning, and inference*, 2nd ed. New York, NY, USA: Cambridge University Press, 2009.
- [37] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., New York, 1969.
- [38] C. K. Kourtellis, "Nonanticipative information theory," Ph.D. thesis, UCY, Nicosia, CY, 2014.
- [39] S. Tatikonda and S. Mitter, "Control over noisy channels," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1196–1201, July 2004.
- [40] C. D. Charalambous and A. Farhadi, "LQG optimality and separation principle for general discrete time partially observed stochastic systems over finite capacity communication channels," *Automatica*, vol. 44, no. 12, pp. 3181–3188, 2008.
- [41] C. D. Charalambous, C. K. Kourtellis, and P. A. Stavrou, "On shannon's duality of a source and a channel and nonanticipative communication and communication for control," in *Coordination Control of Distributed Systems*, ser. Lecture Notes in Control and Information Sciences, J. H. van Schuppen and T. Villa, Eds. Springer International Publishing, 2015, vol. 456, pp. 291–305.
- [42] M. Gastpar, "To code or not to code," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne (E.P.F.L), 2002.
- [43] T. Weissman and N. Merhav, "On causal source codes with side information," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 4003–4013, 2005.
- [44] R. Gray, "Information rates of stationary ergodic finite-alphabet sources," *IEEE Transactions on Information Theory*, vol. 17, no. 5, pp. 516–523, 1971.
- [45] J. D. Deuschel and D. W. Stroock, *Large Deviations*. San Diego: Academic Press, Inc., 1989.
- [46] C. D. Charalambous, S. M. Djouadi, and S. Z. Denic, "Stochastic power control for wireless networks via sdes: probabilistic QoS measures," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4396–4401, Dec 2005.
- [47] C. K. Kourtellis, C. D. Charalambous, and J. J. Boutros, "Nonanticipative transmission for sources and channels with memory," in *IEEE International Symposium on Information Theory (ISIT) (to appear)*, June 14–19 2015.
- [48] P. W. Glynn and D. Ormoneit, "Hoeffding's inequality for uniformly ergodic markov chains," *Statistics & Probability Letters*, vol. 56, no. 2, pp. 143 – 146, 2002.
- [49] C. A. León and F. Perron, "Optimal hoeffding bounds for discrete reversible markov chains," *The Annals of Applied Probability*, vol. 14, no. 2, pp. 958–970, May 2004.
- [50] F. P. Kelly, *Reversibility and Stochastic Networks*. New York, NY, USA: Cambridge University Press, 2011.
- [51] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. Springer-Verlag Inc., New York, 1998.
- [52] P. A. Stavrou, C. K. Kourtellis, and C. D. Charalambous, "Information nonanticipative rate distortion function and its applications," *CoRR*, vol. abs/1405.1593, 2014. [Online]. Available: <http://arxiv.org/abs/1405.1593>
- [53] N. Ma and P. Ishwar, "On delayed sequential coding of correlated sources," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3763–3782, 2011.